

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 09:58:21 ; Search time 1807.19 Seconds

(without alignments)
-740.686 Million cell updates/sec

Title: US-09-501-714-2

Sequence: 1 TCTCACCAGGACTCGGACT.....TTTATTTCATATGCAAGT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 882769 segs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenBml:*

- 1: gb_ba1:*
- 2: gb_ba2:*
- 3: gb_om:*
- 4: gb_ov:*
- 5: gb_pat:*
- 6: gb_ph:*
- 7: gb_pl1:*
- 8: gb_pl2:*
- 9: gb_pr1:*
- 10: gb_pr2:*
- 11: gb_pr3:*
- 12: gb_ro:*
- 13: gb_sts:*
- 14: gb_sy:*
- 15: gb_un:*
- 16: gb_vl:*
- 17: em_fun:*
- 18: em_hum1:*
- 19: em_hum2:*
- 20: em_in:*
- 21: em_om:*
- 22: em_or:*
- 23: em_ov:*
- 24: em_pat:*
- 25: em_ph:*
- 26: em_pl:*
- 27: em_ro:*
- 28: em_sts:*
- 29: em_sy:*
- 30: em_un:*
- 31: em_vl:*
- 32: gb_htg1:*
- 33: gb_htg2:*
- 34: gb_in1:*
- 35: gb_in2:*
- 36: em_ba1:*
- 37: em_ba2:*
- 38: em_hum3:*
- 39: em_hum4:*
- 40: gb_pr4:*
- 41: gb_htg3:*
- 42: gb_htg4:*
- 43: gb_htg5:*
- 44: gb_htg6:*

45: gb_htg7:*

46: em_htg1:*

47: em_htg2:*

48: em_htg3:*

49: em_hum5:*

50: gb_pl3:*

51: gb_pr5:*

52: gb_htg8:*

53: gb_htg9:*

54: gb_htg10:*

55: gb_htg11:*

56: gb_htg12:*

57: gb_htg13:*

58: gb_htg14:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1328	96.5	1701	9 AB028859	AB028859 Homo sapi
2	1150.6	83.6	1159	11 HSA250137	AJ250137 Homo sapi
3	605.2	44.0	745	5 A95223	A95223 Sequence 20
4	323	23.5	89428	43 AC017242	AC017242 Drosophila
5	323	23.5	104278	35 AC005269	AC005269 Drosophila
6	263.2	19.1	145017	57 AC016926	AC016926 Homo sapi
7	171	12.4	145017	57 AC016926	AC016926 Homo sapi
8	113	8.2	332	13 HSA1842A5	Z52396 H. sapiens (
9	101.4	7.4	2656	11 AF061749	AF061749 Homo sapi
10	91.8	6.7	1355	35 AF063011	AF063011 Dictyoste
11	91.4	6.6	975	9 AF080569	AF080569 Homo sapi
12	91.4	6.6	1485	9 AB014888	AB014888 Homo sapi
13	91.4	6.6	1487	9 AB015799	AB015799 Homo sapi
14	91.4	6.6	1489	40 AF075601	AF075601 Homo sapi
15	91.4	6.6	1537	40 AF060703	AF060703 Homo sapi
16	89.8	6.6	2469	9 AB015798	AB015798 Homo sapi
17	88.2	6.5	215297	32 AP001000	AP001000 Homo sapi
18	88.2	6.4	4776	1 AB018706	AB018706 Nitrosomo
19	83.4	6.1	183167	41 AC009546	AC009546 Homo sapi
20	81.2	5.9	1539	12 AB028854	AB028854 Mus muscu
21	81.2	5.9	1546	12 AF035962	AF035962 Mus muscu
22	80.2	5.8	161143	55 AC022634	AC022634 Homo sapi
23	79.6	5.8	1870	35 DMU34904	U34904 Drosophila
24	79.6	5.8	46279	42 AC014815	AC014815 Drosophila
25	79.6	5.8	112886	56 AC010661	AC010661 Drosophila
26	79.2	5.8	4874	2 AF106835	AF106835 Methylovo
27	78.8	5.7	3899	1 STD58360	STD58360 Salmonella
28	78.6	5.7	6201	20 DVTIDNOT	Y07700 Drosophila
29	78	5.7	7412	4 FSCCSC516	M99327 Torpedo cal
30	77.8	5.7	9693	12 AF092536	AF092536 Mus muscu
31	77.8	5.7	124230	12 AC005259	AC005259 Mouse BAC
32	77.6	5.6	28963	43 AC014792	AC014792 Drosophila
33	77.6	5.6	131410	43 AC008319	AC008319 Drosophila
34	77.2	5.6	1218	14 AF088983	AF088983 Mus muscu
35	76.4	5.6	1436	7 AB029547	AB029547 Schizosac
36	76.2	5.5	1140	2 AF053962	AF053962 Campyloba
37	76.2	5.5	1316	1 MAJ10152	AJ010152 Methanosa
38	75.8	5.5	282183	2 CJI1168X4	AL139077 Campyloba
39	75.4	5.5	660	8 CNS01AWK	AL139077 Botrytis
40	75.4	5.5	720	40 CNS01AWK	AL113484 Botrytis
41	75.4	5.5	1580	7 ATDMAJLTK	AL113038 Botrytis
42	75.4	5.5	12029	35 AE001405	AE001405 Plasmidia
43	74.2	5.4	4139	2 AF055368	AF055368 Vibrio ha
44	73.4	5.3	18255	42 AC013054	AC013054 Drosophila
45	73.4	5.3	82393	34 AC003923	AC003923 Drosophila

ALIGNMENTS

```

RESULT 1
LOCUS AB028859 1701 bp mRNA PRI 14-DEC-1999
DEFINITION Homo sapiens mRNA for hdj9, complete cds.
ACCESSION AB028859
VERSION AB028859.1 GI:567165
KEYWORDS hdj9.
SOURCE Homo sapiens cDNA to mRNA, clone_1lb:Stratagene pancreas.
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
          Euthera; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (sites)
          Ohtsuka, K. and Hata, M.
          Mammalian HSP40/DNAJ homologs: Cloning of novel cDNAs and proposal
          for the classification and nomenclature
          Unpublished (1999)
          2 (bases 1 to 1701)
          Ohtsuka, K.
          Direct Submission
          Submitted (10-JUN-1999) to the DDBJ/EMBL/Genbank databases, Kenzo
          Ohtsuka, Aichi Cancer Center Research Institute, Laboratory of
          Experimental Radiology; 1-1 Kanokond, Chikusa-ku, Nagoya, Aichi
          464-8681, Japan (E-mail: khtsuka@aichi-cc.pref.aichi.jp,
          Tel: 81-52-762-6111 (ex. 8845), Fax: 81-52-763-5233)
          Location/Qualifiers
            1. 1701
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone_1lb="Stratagene pancreas"
              240..1316
              /gene="hdj9"
              240..1316
              /gene="hdj9"
              /note="Deduced amino acid sequence of hdj9 is highly
              homologous to C. elegans putative dnaJ protein (247356,
              gene T15H9.1); DnaJ homolog; N-terminal signal peptide
              bp240-305"
              /codon_start=1
              /product="hdj9"
              /protein_id="BAA88307.1"
              /db_xref="GI:567165"
              /translation="MAPONTSPFCLILLYLIGAVIAGRDFYILVPSASIKDKKA
              YRKIAQLHPDNPDPQAEKFDQIGATVLSSEKRODYEGELKDHQSSA
              GTFSHFEDGFMEGTGTRQODRNI PRSDIIVLEVTLEEVAGNEVVRNRPVA
              ROAPGRKRCNQEMKTTQLGPRQOMQEVYCDPCPNKLVNERITVEIEPEVRD
              GMEYPIEGEBPHVDGEPDGLFRIRIKVNHPIFERRGDLYNTVITSLVESLVGEMD
              ITHLDGHKVIHSRDKITRPGLKMKKGEGLPFEDNNNKSLIIFVDVDFPEQLTEE
              AREGIKOLKQSGVKVYNGIGY"
              306..515
              /misc_feature
              /gene="hdj9"
              /note="J domain"
              489 a 321 c 471 g 420 t
BASE COUNT
ORIGIN
Query Match 96.5%; Score 1328; DB 9; Length 1701;
Best Local Similarity 99.1%; Pred. No. 0;
Matches 1334; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

```

```

Db 255 CTGAGCACCTTTGGCCGTGGCTGCTATACCTCATCGGGCGGTGATGGCCGACAGAGAT 314
Qy 271 TTCTATAGATCTTGGGGGTGCTCGAAGTGCCTTAAGATATTAAGAGGCTAT 330
Db 315 TTCTATAGATCTTGGGGGTGCTCGAAGTGCCTTAAGATATTAAGAGGCTAT 374
Qy 331 AGCAACTAGCCCTGACGCTTCATCCGACCGGACCTGATGATCACAAGCCCGAGAG 390
Db 375 AGCAACTAGCCCTGACGCTTCATCCGACCGGACCTGATGATCACAAGCCCGAGAG 434
Qy 391 AAATTCAGATCGGGGTGCTTATAGATCTTCAATAGTATAGTAAAGGAAACAG 450
Db 435 AAATTCAGATCGGGGTGCTTATAGATCTTCAATAGTATAGTAAAGGAAACAG 494
Qy 451 TACGATACCTTATGATGAAGAGATTAAGATGTCATCAGAGCTCCCATGAGACATT 510
Db 495 TACGATACCTTATGATGAAGAGATTAAGATGTCATCAGAGCTCCCATGAGACATT 554
Qy 511 TTTTCACACTCTTTGGGGATTTTGGTTTCATGTTTGAAGAAACCCCTGTCAGACAG 570
Db 555 TTTTCACACTCTTTGGGGATTTTGGTTTCATGTTTGAAGAAACCCCTGTCAGACAG 614
Qy 571 AGAATATTCAGAGAGAGATATTTATTTAGATCTAAGATCACTTTGGAAGAGTA 630
Db 615 AGAATATTCAGAGAGAGATATTTATTTAGATCTAAGATCACTTTGGAAGAGTA 674
Qy 631 TATGACAGAAATTTTGTGAAGTAGTTAGAAACAAACCTGTGCAAGAGAGCTCTGCG 690
Db 675 TATGACAGAAATTTTGTGAAGTAGTTAGAAACAAACCTGTGCAAGAGAGCTCTGCG 734
Qy 691 AAACGAGATGCAATTTGTGGAAGAGATGCGACACCCAGCTGGCCCTGGCGCTTC 750
Db 735 AAACGAGATGCAATTTGTGGAAGAGATGCGACACCCAGCTGGCCCTGGCGCTTC 794
Qy 751 CAAATGACCCAGAGAGTGTGTCGACAGATGCTTAATGTCAAATGTCAGATGAAGAA 810
Db 795 CAAATGACCCAGAGAGTGTGTCGACAGATGCTTAATGTCAAATGTCAGATGAAGAA 854
Qy 811 CCAAGCTGGAAGTAGAATTAAGCTGCGGTGAGACAGCGGATGAGTACCCCTTAT 870
Db 855 CCAAGCTGGAAGTAGAATTAAGCTGCGGTGAGACAGCGGATGAGTACCCCTTAT 914
Qy 871 GGAGAAGTGAAGCTCAGCTGATGAGGAGCTGAGATTAAGCTTCCGAATCAAT 930
Db 915 GGAGAAGTGAAGCTCAGCTGATGAGGAGCTGAGATTAAGCTTCCGAATCAAT 974
Qy 931 GTCAAGCAACCAATATTTGAAGAGAGAGATTTGTACACAATGTGACAGTCTCA 990
Db 975 GTCAAGCAACCAATATTTGAAGAGAGAGATTTGTACACAATGTGACAGTCTCA 1034
Qy 991 TTAGTTGATGCTGCTGGTGGCTTGAATGATATTAATCTACTGATGTCACAAAGTA 1050
Db 1035 TTAGTTGATGCTGCTGGTGGCTTGAATGATATTAATCTACTGATGTCACAAAGTA 1094
Qy 1051 CATATTTCCGGGATTAAGATCAGAGGCTCAGAGCGAANTANTGAAGAAAGGGAAGG 1110
Db 1095 CATATTTCCGGGATTAAGATCAGAGGCTCAGAGCGAANTANTGAAGAAAGGGAAGG 1154
Qy 1111 CTCGCCAATTTGACACACAATATCAAGGCTCTTGTATATCACTTTTGTATGTGAT 1170
Db 1155 CTCGCCAATTTGACACACAATATCAAGGCTCTTGTATATCACTTTTGTATGTGAT 1214
Qy 1171 TTTCACAAAGAAAGTAAACAGAGAGAGCGAGAAAGTAAACAGTAAACAA 1230
Db 1215 TTTCACAAAGAAAGTAAACAGAGAGAGCGAGAAAGTAAACAGTAAACAA 1274
Qy 1231 GGGTCAGTGAAGAGATATCAATGAAGTGAAGATTTGAGAGTAAATTAATTTGAC 1290
Db 1275 GGGTCAGTGAAGAGATATCAATGAAGTGAAGATTTGAGAGTAAATTAATTTGAC 1334
Qy 1291 TTGTTTAAATTAAGTAAATTAAGATATTTATATCTGCAAGGTTTGTGTGTGT 1350

```

Db 1335 TTGTGTTAAATAAGTAAAGCATATTTATATCTGCAAGGTTTTTGTGTGCT 1394
QY 1351 TTTGTTTTTATTTCAATATGCAAGT 1376
Db 1395 TTTGTTTTTATTTCAATATGCAAGT 1420

RESULT 2
HSA250137 1159 bp mRNA PRI 07-JAN-2000
LOCUS HSA250137 1159 bp mRNA
DEFINITION Homo sapiens mRNA for ERJ3 protein (ERJ3 gene).
ACCESSION AJ250137
VERSION AJ250137.1 GI:6688202
KEYWORDS ERJ3 gene; ERJ3 protein.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1159)
Bies, C., Blum, R., Dudek, J. and Zimmermann, R.
Characterization of a mammalian Scjlp homolog
Unpublished
2 (bases 1 to 1159)
Zimmermann, R.
Direct Submission
Submitted (11-OCT-1999) Zimmermann R., Medizinische Biochemie und
Molekularbiologie, Universitaet des Saarlandes, Gebaeude 44,
D-66421 Homburg / Saar, GERMANY
Location/Qualifiers
1. 1159
/organism="Homo sapiens"
/db_xref="taxon:9606"
/tissue_type="placenta"
21. 86
/gene="ERJ3"
/evidence="experimental"
21. 1097
/note="homolog of yeast Scjlp"
/function="putative role in cellular protein folding"
/codon_start=1
/evidence="experimental"
/product="ERJ3 protein"
/protein_id="CAB65118.1"
/db_xref="GI:6688203"
/translation="MAPQNIISFCLILLYLIGAVIAGRDYKILGYPRSAISIKIRKA
YKRLALPHDNPDDPOAEKFODLGAAYEVLSDEKROYDTYGEGLKDGHOSSH
GLFSHFEGDFGFMGGTRRODRNIPRSDIIVDLVLEEVYAGNFYVYKRPYA
RQAPGRKRCNKQEMRTIOLGRRFQMTBVCDECPNKLVAEETLEVEIIEPYRD
GMEYPTIGSEPHVDEPDLRIRKIVNHPLEPRGGDLTYNVTLSVLSLVGFEMD
IYHLDHAKHVISDKITRPAKIMKKEGLPNPDNNINIKSLIITDVSDFPKQLTEE
AREGIKQLKQSGVYKYNLOGY"
21. 1097
/gene="ERJ3"
87. 1094
/gene="ERJ3"
/note="homolog of yeast Scjlp"
/evidence="experimental"
/product="ERJ3"

BASE COUNT 355 a 216 c 314 g 274 t
ORIGIN

Query Match 83.6%; Score 1150.6; DB 11; Length 1159;
Best local similarity 99.5%; Pred. No. 1.8e-258;
Matches 1153; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 176 GACCCGGGACAGAGAACATGGCTCCGCAAGAACTGAGACCTTTGGCTGTGCTGC 235
Db 1 GAACCCGGGACAGAGAACATGGCTCCGCAAGAACTGAGACCTTTGGCTGTGCTGC 60

QY 236 TATACGCTATCGGGGGGCGGTATGCGGAGAGATTCTTAAGATCTTGSGGGGTGCTC 295
|||||

Db 61 TATACGCTATCGGGGGGCGGTATGCGGAGAGATTCTTAAGATCTTGSGGGGTGCTC 120
QY 296 GAATGCCCTCTATTAAGAGATTTAAAAAGCCTATAGAAACTAGCCCTGACGCTATC 355
Db 121 GAATGCCCTCTATTAAGAGATTTAAAAAGCCTATAGAAAACCTAGCCCTGACGCTATC 180
QY 356 CCGACCGGAACCCCGATGATCCACAAGCCCAAGGAGAAATCCAGATCTGGGTCTGCTT 415
Db 181 CCGACCGGAACCCCGATGATCCACAAGCCCAAGGAGAAATCCAGATCTGGGTCTGCTT 240
QY 416 ATGAGTTCTGTCTCAGATAGTGAAGAAAGGAAACAGTACTTATGTGAAGAGAT 475
Db 241 ATGAGTTCTGTCTCAGATAGTGAAGAAAGGAAACAGTACTTATGTGAAGAGAT 300
QY 476 TAAAGATGGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTGGGATTTTG 535
Db 301 TAAAGATGGTCATCAGAGCTCCCATGAGACATTTTTCACACTCTTGGGATTTTG 360
QY 536 GTTTCATGTTTGGAGAAACCCCTGTCAGCAAGACAGAAATTTCCAAAGAGAGATGATA 595
Db 361 GTTTCATGTTTGGAGAAACCCCTGTCAGCAAGACAGAAATTTCCAAAGAGAGATGATA 420
QY 596 TTATTTAGATCTGAAGTCACTTTGGAAGAGATATGACAGAAATTTTGTGGAAGTAG 655
Db 421 TTATTTAGATCTGAAGTCACTTTGGAAGAGATATGACAGAAATTTTGTGGAAGTAG 480
QY 656 TTAGAAACAAACCTGTGGCAAGGCGCTCCGTGCAAAAGGAAATGCAATTTGCGCAAG 715
Db 481 TTAGAAACAAACCTGTGGCAAGGCGCTCCGTGCAAAAGGAAATGCAATTTGCGCAAG 540
QY 716 AGATGCGGACCAACCCAGCTGGGCGCTTCCAAATGACCAAGAGAGTGGTCTGCG 775
Db 541 AGATGCGGACCAACCCAGCTGGGCGCTTCCAAATGACCAAGAGAGTGGTCTGCG 600
QY 776 ACGAATGCCCTTAATGTCAAACTAGTGAATGAAGAACACGCTGGAAGTGAATAAGAGC 835
Db 601 ACGAATGCCCTTAATGTCAAACTAGTGAATGAAGAACACGCTGGAAGTGAATAAGAGC 660
QY 836 CTGGGGTGAGAGACGCGATGGATACCCCTTATTTGGAAGAGTGGACCTGACGTGATG 895
Db 661 CTGGGGTGAGAGACGCGATGGATACCCCTTATTTGGAAGAGTGGACCTGACGTGATG 720
QY 896 GGGAGCCTGGAGATTTTACGGTTCGGATTCAAAGTTGCAACGACCAATATTTAAAGGA 955
Db 721 GGGAGCCTGGAGATTTTACGGTTCGGATTCAAAGTTGCAACGACCAATATTTAAAGGA 780
QY 956 GAGAGATGATTTTACAAATGTCAGACGTCATAGTTGACTGACTGGTGGCTTTG 1015
Db 781 GAGAGATGATTTTACAAATGTCAGACGTCATAGTTGACTGACTGGTGGCTTTG 840
QY 1016 AGATGATATTTACTACCTTGGATGTCACAAGTATATTTCCGGGATTAAGATCACA 1075
Db 841 AGATGATATTTACTACCTTGGATGTCACAAGTATATTTCCGGGATTAAGATCACA 900
QY 1076 GGGCAGAGCAANTANTGCAAGAAAGGAGGCGTCCCACTTTGACACAACATA 1135
Db 901 GGGCAGAGCAANTANTGCAAGAAAGGAGGCGTCCCACTTTGACACAACATA 960
QY 1136 TCAAGGCTCTTTGATATATCACTTTGTGATGAGATTTTCCAAAGACAGTATACAGAG 1195
Db 961 TCAAGGCTCTTTGATATATCACTTTGTGATGAGATTTTCCAAAGACAGTATACAGAG 1020
QY 1196 AAGCAGAGAGATATCAACAGCTACTGAAACAGGGTCAAGTCCAGAGGATATACATG 1255
Db 1021 AAGCAGAGAGATATCAACAGCTACTGAAACAGGGTCAAGTCCAGAGGATATACATG 1080
QY 1256 GACTGCAAGATATTTGAGAGATGAAATTAATTTGACTTTGTTAAATTAAGTAAAGCG 1315
Db 1081 GACTGCAAGATATTTGAGAGATGAAATTAATTTGACTTTGTTAAATTAAGTAAAGCG 1140
QY 1316 ATATTATATATCTGCAAGG 1334
Db 1141 ATATTATATCTGCAAGG 1159

```

RESULT 3
LOCUS A95223
DEFINITION Sequence 20 from Patent WO928461.
ACCESSION A95223
VERSION A95223.1 GI:6779297
KEYWORDS
SOURCE .
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 745)
AUTHORS Danen-Van, O.A. and Noteborn, M.H.
TITLE METHODS AND MEANS FOR INDUCING APOPTOSIS BY INTERFERING WITH
JOURNAL BIP-LIKE PROTEINS
LEAD B V (NL); DANEN VAN OORSCHOT ASTRID ADRI (NL)
FEATURES
Source 1..745
Location/Qualifiers
BASE COUNT 237 a 128 c 184 g 177 t 19 others
ORIGIN
Query Match 44.0%; Score 605.2; DB 5; Length 745;
Best Local Similarity 89.6%; Pred. No. 3, 6e-131;
Matches 660; Conservative 0; Mismatches 74; Indels 3; Gaps 2;

QY 590 GTGATATTATTGATCTGGAAGTCACCTTGGAGAATATATGAGAAATTTGTGG 649
DB 1 GTATATTATTGATGATCTGGAAGTCACCTTGGAGAATATATGAGAAATTTGTGG 60
QY 650 AAGTAGTTAGAAACCAACCTGTGGCAAGGAGGCTCTGTGCAACGGAATGCAATTTGTC 709
DB 61 AAGTAGTTAGAAACCAACCTGTGGCAAGGAGGCTCTGTGCAACGGAATGCAATTTGTC 120
QY 710 GGCAGAAGATGCGGACCAACCACTGGGCTTGGGCTTCCAAATGACCCAGAGGTGG 769
DB 121 GGCAGAAGATGCGGACCAACCACTGGGCTTGGGCTTCCAAATGACCCAGAGGTGG 180
QY 770 TCTGGCAACGAATCCCTTAATGTCAAACTAGTGAATGAGAAGACGCTGGAAATGAAA 829
DB 181 TCTGGCAACGAATCCCTTAATGTCAAACTAGTGAATGAGAAGACGCTGGAAATGAAA 240
QY 830 TAGAGCTGGGCTGAGAGAGCGGATGAGATACCCCTTTATTGAGAAGGTGAGCTCAG 889
DB 241 TAGAGCTGGGCTGAGAGAGCGGATGAGATACCCCTTTATTGAGAAGGTGAGCTCAG 300
QY 890 TGGATGGGAGCCTGGAGATTTACGTTCCGAATCAAAAGTTGTCAAGCAACCAATATTGG 949
DB 301 TGGATGGGAGCCTGGAGATTTACGTTCCGAATCAAAAGTTGTCAAGCAACCAATATTGG 360
QY 950 AAAGGAGAGAGATGATTTGTACAAATGTGACATCTCATTAAGTTGAGTCACTGGTG 1009
DB 361 AAAGGAGAGAGATGATTTGTACAAATGTGACATCTCATTAAGTTGAGTCACTGGTG 420
QY 1010 GCTTTGAGATGATATTACTCACTTGTGATGTCACAAAGTACATATTCCCGGAT -AAG 1068
DB 421 GCTTTGAGATGATATTACTCACTTGTGATGTCACAAAGTACATATTCCCGGAT -AAG 480
QY 1069 ATTCACGAGCCAGAGAGCGGAATNTGGAAGAAAGGGGCTCCCAACTTTGCAAC 1128
DB 481 ATTCACGAGCCAGAGAGCGGAATNTGGAAGAAAGGGGCTCCCAACTTTGCAAC 540
QY 1129 AACCAATATCAAGGCTCTTGTATATCACTTTTGTGATGTGATTTTCCAAAGACAGTTA 1188
DB 541 AACCAATATCAAGGCTCTTGTATATCACTTTTGTGATGTGATTTTCCAAAGACAGTT 600
QY 1189 ACAGAGAGAGCGAGAGAGATCAACAGCTACTGAAACAAGGGTCACTGAGAGAGTA 1248
DB 601 ACAGAGAGAGCGAGAGAGATCAACAGCTACTGAAACAAGGGTCACTGAGAGAGTA 658

```

```

QY 1249 TACATGACTGCAAGAGATTTGAGAGTGAATTAATTTGACTTTGTTAAATTAAGTGA 1308
DB 659 TNCAAATGACCGCAAAANTTTGAAANTGAATTAATTTGNCNTTGTAAATAATTAATTA 718
QY 1309 ATAGAGATATTTATTTA 1325
DB 719 NCCATATTTATNTANTCA 735

RESULT 4
LOCUS AC017242/c
DEFINITION Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***, In ordered
pieces.
ACCESSION AC017242
VERSION AC017242.1 GI:6553744
KEYWORDS HTG; HTGS_PHASE2.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 89428)
Adams, M. and Venter, J.C.
Direct Submission
Submitted (09-DEC-1999) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT
This sequence was identified as CDM:1020998 by the submitter.
For more information on this record e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
Source 1..89428
Location/Qualifiers
BASE COUNT 24712 a 19587 c 20235 g 24894 t
ORIGIN
Query Match 23.5%; Score 323; DB 43; Length 89428;
Best Local Similarity 59.4%; Pred. No. 3, 7e-65;
Matches 586; Conservative 0; Mismatches 392; Indels 9; Gaps 2;

QY 226 CTTGCTCTCTTACCTCACTCGGGCGGTGATTCGCGGACGATTTCTATTAAGATCTTG 285
DB 1420 CAGCTGTCCCTCTTCTGTGTGAGAGATCATTTGCGCGGCGGACTTCTACAAATACTG 1361
QY 286 GGGTGCCCTCGAAGTCCCTTAATTAAGATTAATTAAGAGGCTTAAGAACTACCCCTG 345
DB 1360 AACGTAAAGAAAAGCCCAACGCAAGCAAGTGAAGAAAGGCTTAAGAGGCTTAAGAG 1301
QY 346 CAGCTTCATCCGCGGAGAACCCCTGATGATCCACAGCCAGAGAAATTCAGAGATCT 405
DB 1300 GAGCTGATCCCGATTAAGAAACAGAGACGCCGATGCTCCACAAATTCAGAGACTG 1241
QY 406 GGTGCTGCTTAAGAGCTTCTGCAATAGTGAAGAAAGGAAACAGTACGATTAATGCT 465
DB 1240 GGAGGGCTTCAAGAGTCCCTTCATCCGCAACAAAGCAAGAGACCTTACGACCGTGG 1181
QY 466 GAAGAAGATT---AAAAGATGTCATCAGAGCTCCATGGAGACATTTTTCACACTTC 522
DB 1180 GAGGAATGCTTCAGAAAGAGAGGCAATGATGATCACGGTGTGATCCGTTCTTACGCTTC 1121
QY 523 TTTGGGATTTTGTGATTTTGTGAGAAACCCCTGTCAGCAAGACGAATATTCCA 582
DB 1120 TTTGGGACTTCGCGCTTCACTTGGGTGATGAGCCAGCAAG-----ATGCTCG 1067
QY 583 AGAGAAATGATTAATTTATGTAGATCTAAGATCTCACTTTGGAAGAAAGTATATGAGAAAT 642
DB 1066 CGAGCGCGGATATCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1007

```



```

OY 643 TTGTGAGAGTACTTAGAAAACAACCTGGGCAAGCGCTCCTGGCAACGAAGTGC 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1006 TTGTGGAATTTGTGAGAACAGCTGTACAAACCCCGCTCAGGACACGAGAAATGC 947
OY 703 AATTGCGGCAAGAGATGGGACACCCAGCTGGGCGCTTCCAAATGACCCAG 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 946 AACTGCCGCGAGAGATGTGACCCGGAACCTTGGACCCGGGGCTTCCAGATGATCCAA 887
OY 763 GAGGTGCTGCGAGCAATGCCCTAATGTCAAACTAGTGAATGAAGACGACGTGAA 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 886 CAGACGCTGCGAGCAAGTGTCCCAAGTGAAGCTAGTCAAGAGAGCGCACATTTGAG 827
OY 823 GTGAATATGAGCTGGGGTGAAGACGCGATGAGTACCCCTTTATGGAAGAGTGAG 882
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 826 ATCGAGGTGGAACAGGCAATGTCGACGGCAAGACGAGGTGCGGGCGAGGGGAG 767
OY 883 CCTCAGCTGAGATGGGAGCGCTGAGATTTACGGTCCGATCAAAAGTTGTCAGCACCCA 942
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 766 CCACATATGATGGGAGCGCCGCGACCTCATTTGTGCGGTTCAACAATGCCGATCCG 707
OY 943 ATATTGAAAGAGAGAGATGATTTGTACCAAAATGTGACAGTCTCATTTAGTTGACTCA 1002
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 706 CGATTCTCGGCAAGATGATGTGTACAGACAGGTGACATGCTGCAAGGATGCC 647
OY 1003 CTGGTTGGCTTTAGATGATATTACTCACTTGGATGTGTACAGAGTACATATTTCCCGG 1062
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 646 CTGTGTGGCTTCTCCATGTGATCAACGACCTTGTGATGACAGCTGGCCGTACAGAGA 587
OY 1063 GATAGATCACCAGCGAGGAGGAANTANTGGAAGAAAGGAGGCGCCCAACTTT 1122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 586 GAGAGAGTAACTGTGCGCCGCGCTAGATCCGCAAGAGGCGAGGCGATGCCAATCTT 527
OY 1123 GACACACAAATATCAAGGCGCTTTGATATCACTTTTGTGATTTTCCAAAATA 1182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 526 GAGAACAAACAACCTCAACGCGCACTGTACATCACTTGATGTGGATGCCCAAGAG 467
OY 1183 CAGTTACAGAGAGACGCGAGAGAGGT 1209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 466 GATCTTACGAGAGAGACAGAGAGGT 440
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AC005269 104278 bp DNA INV 07-JUL-1998
LOCUS Drosophila melanogaster DNA sequence (Pis DS00764 (D273) and
DEFINITION DS00501 (D274)), complete sequence.
AC005269 AC004571 AC004572
AC005269.1 GI:3293205
HTG.
Drosophila melanogaster (Subclones in tet from P1 clones DS00764
(D273) and DS00501 (D274)) DNA.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 104278)
Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Abpayani,A., Arcinas,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Humastil,S.R., Katra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacle,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.
Sequencing of Drosophila chromosome, region 21C3-21C7
Unpublished (1997)
2 (bases 1 to 104278)
Celiker,S.E., George,R.A., Galle,R.F., Hoskins,R.A.,
Svirskas,R.R., Harris,N.L., Abpayani,A., Arcinas,T.T., Baxter,E.,
Blazej,R.G., Chavez,C., Chew,M., Doyle,C.M., Farfan,D.E.,
Flanagan,J., Houston,K.A., Humastil,S.R., Katra,K., Kearney,L.,
Kim,S.H., Lee,B., Lomotan,M.A., Mak,J., Mazda,P., Mok,M.S.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacle,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.

```

```

TITLE
JOURNAL
Moshrefi,A.R., Moshrefi,M., Nixon,K., Pacle,J.M., Park,S.,
Pfeiffer,B., Punch,E., Snir,E., Twomey,B., Wan,K.H., Whitelaw,K.R.,
Yee,A., Zhang,R., Zieran,L.L. and Kimmel,B.E.
Direct Submission
Submitted (07-JUL-1998) Berkeley Drosophila Genome Project, MS
64-121, Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, US
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory, MS 64-121
Berkeley, CA 94720
For further information about this sequence, including its location
and relationship to other sequences, please visit our sequence
archive web site (http://fruitfly.berkeley.edu/sequence/) or send
email to drosophila@mcg.lbl.gov.
Library locations: 135-8, 21-6.
Location/Qualifiers
1. 104278
FEATURES
source
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="2L"
/map="21C3-21C7"
/clone="Pis DS00764 (D273) and DS00501 (D274)"
/note="DS00764 (d273) extends from bp 1 to bp 38,674 and
DS00501 (d274) extends from bp 20,505 to bp 104,278."
BASE COUNT 28186 a 23704 c 23278 g 29110 t
ORIGIN
Query Match 23.5%; Score 323; DB 35; Length 104278;
Best Local Similarity 59.4%; Pred. No. 3.7e-65;
Matches 586; Conservative 0; Mismatches 392; Indels 9; Gaps 2;
OY 226 CTGTGCTGTATACATCATCGGGCGGTGATGTCGCGAGATTTCTAATGATCTG 285
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16821 CAGCTGTCCCTCTGCGTGGAGAGTCAATTTCCGCGGGAGCTCTCAAAATACTG 16880
OY 286 GGGGTGCTGAGAGTGCCTCTATTAAGGATTTAAAAAGGCTTATAGAACTAGCCCTG 345
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16881 AACTTAAGAAAACCAACACGACGAAAGTGAAGAGTGATCCCGCTTGGCCAG 16940
OY 346 CAGCTTATCCCGACCGAACCCTGATGATCCACAAACCCAGGAATTCAGAGTCTG 405
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 16941 GAGCTGATCCGATTAAGAACAGAGACCGGATGCTTCACAAATTCAGAGACTG 17000
OY 406 GTGCTGCTTATAGAGTCTGTGATGATGAGAAAGCAAGTACGATCTATGATG 465
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17001 GAGCGGCTACGAAAGTCTCTCCATCCGAGCAAAAGGAAAGTACGACCGTGGCGC 17060
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 466 GAAGAGGATTT--AAAAGTGTATCAGAGCTCCCATGAGACATTTTTCACACTTC 522
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17061 GAGGAATGCTTCAAGAGAGGAGGATGATGATCAGGTGTGATCGTCTCTAGCTTC 17120
OY 523 TTGGGAGATTTTGTTCATGTTTGGAGAAACCCCTGTAGCAAGACAGAAATATTC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17121 TTGGGAGATTTTGTTCATGTTTGGAGAAACCCCTGTAGCAAGACAGAAATATTC 582
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 583 AGAAGAGTATATTTATGATCTAGATCTAGATCTTGGAGAAAGATATGCAAGAT 642
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17175 CGAGCGCCGATATCGTAATGAGACTTGTACCTTCCCTGGAGAGCTATATCTCGGAAAC 17234
OY 643 TTGTGGAAGTATTAGAAACAACCTGTGGCAAGGAGGCTCTCGCAAAACGAGAGTGC 702
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17235 TTGTGGAATTTGTGAGAGCAACGCTGTAAAGAAACCCGCTTCAGACCAAGAAATGC 17294
OY 703 AATTGTGGCAAGAGATGCGGACACCACTGGGCGCTTGGCGCTTCAATAGACCAAG 762
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17295 AACTGCCGCGAGAGATGTGACCCGGAACCTTGGAGACCGGGGCTTCCAGATGATCCAA 17354
OY 763 GAGGTGCTGCGAGCAATGCCCTAATGTCAAACTAGTGAATGAAGAACAGAGTGGAA 822
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 17355 CAGAGGTGTGCGAGAGTGTCCCAACGTAAGCTAGTCAACGAGAGGCGCACATTTGGAG 17414

```

QY 823 CTAGAAATAGACCTGGGGTGAGACGCGCATGTACCCCTTTATTGAGAAAGTGCAG 882
 Db 17415 ATCAGAGGTGAACAGAGGCATGTCGACGGCCAGAGACGAGGTTCGTGCGGAGGAG 17474
 QY 883 CCTACGCGATGGGAGCCCTGGAGATTTCAGGTTCCGAATCAAAGTTGTCACACCCA 942
 Db 17475 CCACATATCGATGGGAGCCCGGCGACCTCATTTGTGGGGTTCAACAAATGCCCATCCG 17534
 QY 943 ATATTGAAAGAGAGAGATGATTGTATACAAATGTGACATCTCATTTAGTTAGTCA 1002
 Db 17535 CGATTCCGCGCAGAAATGATGATCTGTACAGCAACGTGACATCATCTGCAGAGATCC 17594
 QY 1003 CTGTGTGGCTTTGAGATGATATTACTTGTGATGATGATGATGATGATGATGATG 1062
 Db 17595 CTGTGTGGCTTTGAGATGATATTACTTGTGATGATGATGATGATGATGATGATG 17654
 QY 1063 GATAAGATTCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1122
 Db 17655 GAGAGAGTAACTGCGCGCGCGCGCTAGAAATCCGCAAGAGGCGGCGCATGCGCAACTTT 17714
 QY 1123 GACAAACAATATCAAGGCGCTTTGTATATCACTTTGATGATGATGATGATGATGATG 1182
 Db 17715 GAGAAACAACACTCACCGGACCTGTACATCACTTCATGATGATGATGATGATGATG 17774
 QY 1183 CAGTTAACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1209
 Db 17775 GATCTTACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 17801

RESULT 6
 AC016926
 LOCUS AC016926 145017 bp DNA HTG 19-FEB-2000
 DEFINITION Homo sapiens clone Rpl1-35G16, *** SEQUENCING IN PROGRESS ***, 60
 unoriented pieces.
 ACCESSION AC016926
 VERSION AC016926.3 GI:7007843
 KEYWORDS HTG: HTGS_PHASE1.
 SOURCE human.
 ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homidae; Homo.
 1 (bases 1 to 145017)

REFERENCE
 AUTHORS Muzny,D.M., Adams,C., Bailey,M., Barbite,J., Blankenburg,K.,
 Boretta,B., Bouck,J., Bowler,S., Brooks,A., Bunney,C., Bunnac,C.,
 Burdett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
 David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
 Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
 Forcum-Tansey,J., Frintz,P., Ganesh,R., Gorrell,J.H., Gorrell,L.L.,
 Guenava,M., Harris,K., Hernandez,J., Hodgson,A., Hogues,M.,
 Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,X., Jones,M.,
 Kelly,S., Kondejewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
 Lichtarge,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
 Lucier,R., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
 Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
 Osvall,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
 Oulles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
 Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugan,R.,
 Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wabnah,M.,
 Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
 Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
 Gibbs,R.
 Direct Submission
 Unpublished
 2 (bases 1 to 145017)
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 COMMENT

Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Feb 19, 2000 this sequence version replaced gi:6671836.
 ----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc_help@bcm.tmc.edu
 ----- Project Information
 Center project name: HMR
 Center clone name: Rpl1-35G16
 ----- Summary Statistics
 Sequencing vector: M13; 108821
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.980611
 Consensus quality: 90208 bases at least Q40
 Consensus quality: 108595 bases at least Q30
 Estimated quality: 11831 bases at least Q20
 Estimated insert size: 145017; agarose-IP estimation
 Estimated insert size: 134603; sum-of-contigs estimation
 Quality coverage: 1.4x in Q20 bases; agarose-IP estimation
 Quality coverage: 1.5x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 60 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 784 783: contig of 783 bp in length
 * 804 803: gap of unknown length
 * 1653 1652: contig of 843 bp in length
 * 1673 1672: gap of unknown length
 * 2517 2516: contig of 844 bp in length
 * 2537 2536: gap of unknown length
 * 3384 3383: contig of 847 bp in length
 * 3404 3403: gap of unknown length
 * 4442 4441: contig of 1038 bp in length
 * 4462 4461: gap of unknown length
 * 5250 5249: contig of 788 bp in length
 * 5270 5269: gap of unknown length
 * 6137 6136: contig of 867 bp in length
 * 6157 6156: gap of unknown length
 * 7654 7653: contig of 1497 bp in length
 * 7674 7673: gap of unknown length
 * 8491 8490: contig of 817 bp in length
 * 8511 8510: gap of unknown length
 * 10429 10428: contig of 1918 bp in length
 * 10449 10448: gap of unknown length
 * 11243 11242: contig of 794 bp in length
 * 11263 11262: gap of unknown length
 * 12563 12562: contig of 1300 bp in length
 * 12583 12582: gap of unknown length
 * 13361 13360: contig of 778 bp in length
 * 13381 13380: gap of unknown length
 * 14316 14315: contig of 935 bp in length
 * 14336 14335: gap of unknown length
 * 14336 14335: contig of 1415 bp in length
 * 15751 15750: gap of unknown length
 * 15771 15770: gap of unknown length
 * 17209 17208: contig of 1438 bp in length
 * 17229 17228: gap of unknown length
 * 18285 18284: contig of 1056 bp in length
 * 18305 18304: gap of unknown length
 * 18305 18304: gap of unknown length
 * 19188 19187: contig of 883 bp in length
 * 19208 19207: gap of unknown length
 * 21137 21136: contig of 1929 bp in length
 * 21157 21156: gap of unknown length
 * 22690 22689: contig of 1533 bp in length
 * 22710 22709: gap of unknown length
 * 25747 25746: contig of 3037 bp in length
 * 25767 25766: gap of unknown length
 * 27267 27266: gap of 1481 bp in length
 * 27268 27267: gap of unknown length
 * 28633 28632: contig of 1365 bp in length
 * 28653 28652: gap of unknown length
 * 30433 30432: contig of 1780 bp in length
 * 30452 30451: gap of unknown length

30453 32234: contig of 1782 bp in length
32235 32254: gap of unknown length
32255 34215: contig of 1961 bp in length
34216 34235: gap of unknown length
34236 35966: contig of 1731 bp in length
35967 35986: gap of unknown length
35987 37169: contig of 1183 bp in length
37170 37189: gap of unknown length
37190 39327: contig of 2138 bp in length
39328 39347: gap of unknown length
39348 40797: contig of 1450 bp in length
40798 40817: gap of unknown length
40818 42978: contig of 2161 bp in length
42979 42998: gap of unknown length
42999 45518: contig of 2520 bp in length
45519 45538: gap of unknown length
45539 47419: contig of 1881 bp in length
47420 47439: gap of unknown length
47440 49427: contig of 1988 bp in length
49428 49447: gap of unknown length
49448 51749: contig of 2302 bp in length
51750 51769: gap of unknown length
51770 53862: contig of 2093 bp in length
53863 53882: gap of unknown length
53883 56866: contig of 2984 bp in length
56867 56886: gap of unknown length
56887 59350: contig of 2464 bp in length
59351 59370: gap of unknown length
59371 61846: contig of 2476 bp in length
61847 61866: gap of unknown length
61867 63930: contig of 2064 bp in length
63931 63950: gap of unknown length
63951 65797: contig of 1847 bp in length
65798 65817: gap of unknown length
65818 69175: contig of 3358 bp in length
69176 69195: gap of unknown length
69196 72751: contig of 3556 bp in length
72752 72772: gap of unknown length
72773 75841: contig of 3070 bp in length
75842 75861: gap of unknown length
75862 78423: contig of 2562 bp in length
78424 78443: gap of unknown length
78444 81402: contig of 2959 bp in length
81403 81422: gap of unknown length
81423 84125: contig of 2703 bp in length
84126 84145: gap of unknown length
84146 87688: contig of 3543 bp in length
87689 87708: gap of unknown length
87709 91413: contig of 3705 bp in length
91414 91433: gap of unknown length
91434 95783: gap of 4350 bp in length
95784 95803: gap of unknown length
95804 98920: contig of 3117 bp in length
98921 98940: gap of unknown length
98941 102954: contig of 4014 bp in length
102955 102974: gap of unknown length
102975 107971: contig of 4997 bp in length
107972 107991: gap of unknown length
107992 111651: contig of 3660 bp in length
111652 111671: gap of unknown length
111672 115722: contig of 4051 bp in length
115723 115742: gap of unknown length
115743 121800: contig of 6058 bp in length
121801 121820: gap of unknown length
121821 127723: contig of 5903 bp in length
127724 127743: gap of unknown length
127744 133403: contig of 5660 bp in length
133404 133423: gap of unknown length
133424 139012: contig of 5589 bp in length
139013 145017: gap of unknown length
145017 145017: contig of 5985 bp in length.

FEATURES
sourceLocation/Qualifiers
1. .145017

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-35G16"
BASE COUNT 40495 a 30419 c 31541 g 41252 t 1310 others
ORIGIN

Query Match 19.1%; Score 263.2; DB: 57; Length 145017;
Best Local Similarity 98.9%; Pred. No. 3.5e-51;
Matches 265; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCTCACCAGGAGTGGGACTCCCGGAAGTGGACCGGAGAGAGGGGGCTAGTACTG 60
Db 128202 TCTCACCAGGAGTGGGACTCCCGGAAGTGGACCGGAGAGAGGGGGCTAGTACTG 128261
QY 61 TCTCTGGCGAGGAGGAGACCCCGCCCGCGGTGTGAGGGCGCCCTCAGAGGCCG 120
Db 128262 TCTCTGGCGAGGAGGAGACCCCGCCCGCGGTGTGAGGGCGCCCTCAGAGGCCG 128321
QY 121 GTGGGCTGGCAGCGCCGCGCGCGCGGAGAGGCTGTGAGAGTGTGTGAACAGACC 180
Db 128322 GTGGGCTGGCAGCGCGCGCGCGGAGAGGCTGTGAGAGTGTGTGAACAGACC 128381
QY 181 CGGAGACAGAGAACCATGCTCCGACAGACCTTTGCTGCTGTCTGTATAC 240
Db 128382 CGGAGACAGAGAACCATGCTCCGACAGACCTTTGCTGCTGTCTGTATAC 128441
QY 241 CTCATCGGGGGGTGATTGCCGAGAC 268
Db 128442 CTCATCGGGGGGTGATTGCCGAGTAC 128469

RESULT 7
AC016926/c DNA 19-FEB-2000
LOCUS
DEFINITION Homo sapiens clone RP11-35G16, *** SEQUENCING IN PROGRESS ***, 60
unordered pieces.

ACCESSION AC016926 GI:7007843
VERSION AC016926.3
KEYWORDS HTG; HTGS; PHASEL.
SOURCE human.
ORGANISM Homo sapiens

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 145017)

Muzny,D.M., Adams,C., Bailey,M., Barberia,J., Blankenburg,K.,
Bodota,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunz,C.,
Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
Forcum-Tansey,J., Frantz,P., Ganesb,R., Gorrell,J.H., Gorrell,L.L.,
Guevarra,W., Harris,K., Hernandez,J., Hodgson,A., Hughes,M.,
Holloway,C., Hosak,H., Jackson,L.E., Jackson,L., Jia,Y., Jones,M.,
Kelly,S., Kondewski,N., Kong,Y., Kovar,C., Leal,B., Li,Z.,
Lichter,O., Liu,J., Liu,W., Logan,O., Lozano,R.J., Lu,J.,
Lucier,K., Martin,R., Martinez,C., McLeod,M.P., Mei,G., Morgan,M.,
Morris,S., Nash,S., Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S.,
Oswal,G., Parish,B., Paxton,S., Payton,B., Perez,L., Pu,L.L.,
Quiles,M., Reiter,D., Rives,M., Samuel,S., Say,J., Scherer,S.,
Shah,E., Shen,H., Simon,M., Sparks,A., Stamps,A., Sugrany,R.,
Taber,P., Taylor,T., Vasquez,L., Vinson,R., Vo,O., Wahbeh,M.,
Watlington,S., Weinstein,G., Weinstein,I.R., Williamson,A.,
Worley,K., Wren,J., Wrenford,G., Yu,W., Zhou,X., Nelson,D. and
Gibbs,R.

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 145017)
AUTHORS Worley,K.C.

COMMENT Direct Submission
Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced g1:6671836.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: HMR
Center clone name: RP11-35G16
----- Summary Statistics
Sequencing vector: M13; 108821
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 90208 bases at least Q40
Consensus quality: 108595 bases at least Q30
Consensus quality: 11831 bases at least Q20
Estimated insert size: 145017; agarose-fp estimation
Estimated insert size: 134603; sum-of-ctgls estimation
Quality coverage: 1.4x in Q20 bases; agarose-fp estimation
Quality coverage: 1.5x in Q20 bases; sum-of-ctgls estimation

* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
784 783: contig of 783 bp in length
804 803: gap of unknown length
1653 1652: contig of 849 bp in length
1673 1672: gap of unknown length
2517 2516: contig of 844 bp in length
2537 2536: gap of unknown length
3383 3383: contig of 847 bp in length
3384 3403: gap of unknown length
3404 4441: contig of 1038 bp in length
4442 4461: gap of unknown length
4462 5249: contig of 788 bp in length
5250 5269: gap of unknown length
5270 6136: contig of 867 bp in length
6137 6156: gap of unknown length
6157 7653: contig of 1497 bp in length
7654 7673: gap of unknown length
7674 8490: contig of 817 bp in length
8491 8510: gap of unknown length
8511 10428: contig of 1918 bp in length
10429 10448: gap of unknown length
10449 11242: contig of 794 bp in length
11243 11262: gap of unknown length
11263 12562: contig of 1300 bp in length
12563 12582: gap of unknown length
12583 13360: contig of 778 bp in length
13361 13380: gap of unknown length
13381 14315: contig of 935 bp in length
14316 14335: gap of unknown length
14336 15750: contig of 1415 bp in length
15751 15770: gap of unknown length
15771 17208: contig of 1438 bp in length
17209 17228: gap of unknown length
17229 18284: contig of 1056 bp in length
18285 18304: gap of unknown length
18305 19187: contig of 883 bp in length
19188 19207: gap of unknown length
19208 21136: contig of 1929 bp in length
21137 21156: gap of unknown length
21157 22689: contig of 1533 bp in length
22690 22709: gap of unknown length
22710 25746: contig of 3037 bp in length
25747 25766: gap of unknown length
25767 27247: contig of 1481 bp in length
27248 27267: gap of unknown length
27268 28632: contig of 1365 bp in length

28633 28652: gap of unknown length
28653 30432: contig of 1780 bp in length
30433 30452: gap of unknown length
30453 32234: contig of 1782 bp in length
32235 32254: gap of unknown length
32255 34215: contig of 1961 bp in length
34216 34235: gap of unknown length
34236 35966: contig of 1731 bp in length
35967 35986: gap of unknown length
35987 37189: contig of 1183 bp in length
37190 37199: gap of unknown length
37199 39327: contig of 2138 bp in length
39328 39347: gap of unknown length
39348 40797: contig of 1450 bp in length
40798 40817: gap of unknown length
40818 42978: contig of 2161 bp in length
42979 42998: gap of unknown length
42999 45518: contig of 2520 bp in length
45519 45538: gap of unknown length
45539 47419: contig of 1881 bp in length
47420 47439: gap of unknown length
47440 49427: contig of 1988 bp in length
49428 49447: gap of unknown length
49448 51749: contig of 2302 bp in length
51750 51769: gap of unknown length
51770 53862: contig of 2093 bp in length
53863 53882: gap of unknown length
53883 56866: contig of 2984 bp in length
56867 56886: gap of unknown length
56887 59350: contig of 2464 bp in length
59351 59370: gap of unknown length
59371 61846: contig of 2476 bp in length
61847 61866: gap of unknown length
61867 63930: contig of 2064 bp in length
63931 63950: gap of unknown length
63951 65797: contig of 1847 bp in length
65798 65817: gap of unknown length
65818 69175: contig of 3358 bp in length
69176 69195: gap of unknown length
69196 72751: contig of 3556 bp in length
72752 72771: gap of unknown length
72772 75841: contig of 3070 bp in length
75842 75861: gap of unknown length
75862 78423: contig of 2562 bp in length
78424 78443: gap of unknown length
78444 81402: contig of 2959 bp in length
81403 81422: gap of unknown length
81423 84125: contig of 2703 bp in length
84126 84145: gap of unknown length
84146 87688: contig of 3543 bp in length
87689 87708: gap of unknown length
87709 91413: contig of 3705 bp in length
91414 91433: gap of unknown length
91434 95783: contig of 4350 bp in length
95784 95803: gap of unknown length
95804 98920: contig of 3117 bp in length
98921 98940: gap of unknown length
98941 102954: contig of 4014 bp in length
102955 102974: gap of unknown length
102975 107971: contig of 4997 bp in length
107972 107991: gap of unknown length
107992 111651: contig of 3660 bp in length
111652 111671: gap of unknown length
111672 115722: contig of 4051 bp in length
115723 115742: gap of unknown length
115743 121800: contig of 6058 bp in length
121801 121820: gap of unknown length
121821 127723: contig of 5903 bp in length
127724 127743: gap of unknown length
127744 133403: contig of 5660 bp in length
133404 133423: gap of unknown length
133424 138012: contig of 5589 bp in length
139013 139032: gap of unknown length

FEATURES	* 139033	145017: contig of 5985 bp in Length.
SOURCE	Location/Qualifiers	
	1. .145017	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="RP11-35G16"	
BASE COUNT	40495 a 30419 c 31541 g 41252 t	1310 others
ORIGIN		
Query Match	12.4%; Score 171; DB 57; Length 145017;	
Best Local Similarity	100.0%; Pred. No. 1.2e-29;	
Matches 171; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
OY 1206	AGGTATCAACACGCTACTGAAACAGGGCTAGTGCAGAGGATATACATGAGCTGCACAG	1265
Db 66241	AGGTATCAACACGCTACTGAAACAGGGCTAGTGCAGAGGATATACATGAGCTGCACAG	66182
OY 1266	ATATTGACAGTCATATAAATTGGACTTTGTTTAAATAAGTCATATGAGATATTATTA	1325
Db 66181	ATATTGACAGTCATATAAATTGGACTTTGTTTAAATAAGTCATATGAGATATTATTA	66122
OY 1326	TCATGACAGTTTATTTGCTGTGTTTGTGTTTATTTTCAATATGCAAGT	1376
Db 66121	TCATGACAGTTTATTTGCTGTGTTTGTGTTTATTTTCAATATGCAAGT	66071
RESULT 8		
HSAL1842A5/c	332 bp	DNA
LOCUS	HSAL1842A5	STS
DEFINITION	H.sapiens (D3S3570) DNA segment containing (CA) repeat; clone	23-MAR-1996
ACCESSION	AF01842a5; single read, sequence tagged site.	
VERSION	252396	
KEYWORDS	252396.1 GI:1233696	
SOURCE	STS; CA repeat; dinucleotide repeat; GT repeat; microsatellite	
ORGANISM	human.	
REFERENCE	Homo sapiens	
TITLE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;	
AUTHORS	Primates; Catarrhini; Homiidae; Homo.	
JOURNAL	1 (bases 1 to 332)	
REFERENCE	Weissenbach, J.	
AUTHORS	Direct Submission	
	Submitted (01-SEP-1995) Genethon, B.P. 60, 91002 Evry Cedex France.	
	E-mail: Jean.Weissenbach@genethon.fr	
	2 (bases 1 to 332)	
	Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A.,	
	Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G.,	
	Morissette, J. and Weissenbach, J.	
	A comprehensive genetic map of the human genome based on 5,264	
	microsatellites	
	Nature 380 (6570), 152-154 (1996)	
	96176476	
	full automatic.	
	Location/Qualifiers	
	1. .332	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="3"	
	/cell_line="CEPH 134702"	
	/clone_lib="genomic DNA"	
	/note="Cloning vector is M13mp18"	
BASE COUNT	98 a 88 c 44 g 90 t	12 others
ORIGIN		
Query Match	8.2%; Score 113; DB 13; Length 332;	
Best Local Similarity	95.0%; Pred. No. 2.8e-16;	
Matches 113; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
OY 1093	TGGAGAAAGGGGAGGCGTCCCAACTTGGACACACAAATATCATCAAGGCTCTTTGATA	1152
Db 327	TGGAGAAAGGGGAGGCGTCCCAACTTGGACACACAAATATCATCAAGGCTCTTTGATA	268

RESULT	9
LOCUS	AF061749
DEFINITION	Homo sapiens 2656 bp mRNA
ACCESSION	AF061749
VERSION	AF061749.1
KEYWORDS	GI:3372676
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	Schilling, B., De-Medina, T., Syken, J., Vidal, M. and Munger, K.
TITLE	1 (bases 1 to 2656)
JOURNAL	A novel human DnaJ protein, hrid-1, a homolog of the Drosophila tumor suppressor protein Tid56, can interact with the human papillomavirus type 16 E7 oncoprotein
MEDLINE	Virology 247 (1), 74-85 (1998)
REFERENCE	98354343
AUTHORS	2 (bases 1 to 2656)
REFERENCE	Schilling, B., De-Medina, T., Syken, J., Vidal, M. and Munger, K.
TITLE	Direct Submission
JOURNAL	Submitted (27-APR-1998) Pathology, Harvard Medical School, 200 Longwood Avenue, Boston, MA 02115-5701, USA
FEATURES	Location/Qualifiers
source	1..2656
gene	/organism="Homo sapiens"
CDS	/db_xref="taxon:9606"
	/chromosome="16"
	/map="16p13.3: between D16S3070 and D16S510"
	1..2656
	/gene="TID1"
	32..1474
	/gene="TID1"
	/note="hrid-1 protein; DnaJ homolog"
	/codon_start=1
	/product="tumorous imaginal discs protein Tid56 homolog"
	/protein_id="AAC29066.1"
	/db_xref="GI:3372677"
	/translation="MAARCSFRLVLYVGFRLPAISRGARPPREGVGAAMLRSKLSVPAAVSSITSCGPRALLILAPRGVSLTGKHPDLFCITASPHISAPLAEDYVOILGVPRNASQREIKKATYQLAKKTHPTDNDPKAKRESQALAEAYLSDEYKRRKYDAYGSGFEDGASGSHSYWKGPTVDPDELFRKIFREFSSSSGEDQTVFDPOETFMELTFRQAAQVKEFTEPVTIMDICEKNGKNGNEKRVQVCHYCGSGGMMTINTGPFVMSRCRGGRSIIISPVCAVCGAGAKOKKRMILPVAGVEDGQVPMVGRKELFIFRVQIKSPFRRDAGDISHDLFTSLAOLILGGFARAQGYELTINVIIPGTONDOKIRMGGSGLIPRINSYGGDHYTHIKRIYVPRKRLTSQOSILSYAEDFIDVEGVNVTLLSSGSGMDSAGSKARREADEDEBGLSKAKKFTS"
BASE COUNT	638 a 694 c 745 g 579 t
ORIGIN	
Query Match	7.4%; Score 101.4; DB: 13; Length 2656;
Best Local Similarity	61.6%; Pred. No. 1.6e-13;
Matches 162; Conservative	0; Mismatches 101; Indels 0; Gaps 0;
238	TACCTCATCGGGGCGGTGTTCCGAGACAGATCTTAAATCTTGGGGGTCCTCGA 297
278	TTCCACAGAGAGGCCCTTTGGCCAAAGAGATTTATTCAGATATTAGAGAGTGCCTGA 337
238	AGTGCCTCTATTAAGATATTAAGAAAGCCTATTAGAAATAGCCCTGACGCTTATCC 357
338	AATGCCAGACAGAAAGAGTCAAGAAAGCCTATTATCAGCTTGCCAAAGATACCCCT 397
358	GACCGGAACCCCTGATGATCCACACAGCCAGAGAAATTCAGAGATCTGGGTCTCTAT 417
398	GACACAAATTAAGATGATCCCAACAGGAGAAAGTTCCTCCACTGGCACAAGCCTAT 457

JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1489)
AUTHORS Zhang, W., Man, T., Yuan, Z. and Cao, X.
TITLE Direct Submission
JOURNAL Submitted (01-JUL-1998) Department of Immunology, Shanghai
Brilliance Biotechnology Institute and Second Military Medical
University, 800 Xiangyin Road, Shanghai 200433, P.R. China

FEATURES
SOURCE location/Qualifiers
1..1489

gene /organism="Homo sapiens"
/db_xref="taxon:9606"
1..1489
CDS /gene="HSJ2"
133..858
/gene="HSJ2"
/note="similar to Mus musculus DNAJ protein encoded by the
sequence presented in Genbank Accession Number U95607, and
to other heat shock proteins"
/product="heat shock J2 protein"
/protein_id="AA04394.1"
/db_xref="GI:5441950"

BASE COUNT 443 a 269 c 337 g 440 t
ORIGIN

Query Match 6.6%; Score 91.4; DB 40; Length 1489;
Best Local Similarity 66.5%; Pred. No. 3.3e-11;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATAAGATCTTGGGGGTCCTCGAAGTCCCTCTATAAGATATTAAAGGCC 327
DB 139 GATTACTATGAAGTTCTGACGCGTGCAAGACATGCCACCCGAGATTTAAAGGCA 198
QY 328 TATAGAACTAGCCCTGACCTTCATCCGACCGGAACCTG--ATGATCCACAAGCC 384
DB 199 TATCGGAACTGGCAGCTGCAAGTGCATCAATTAATCTGAGATTAAGAAAGCA 258
QY 385 CAGAGAAATTCAGATCTGGGTGCTCTATGAGGTTCTGATAGTGAAGAACGG 444
DB 259 GAGAGAAATTCAGAGCAAGTAGCGAGGCAATGAGTGTGCGATGCTTAAGAACGG 318
QY 445 AAACAGTACGATCTTATGCTGAAGAGGATTTAAAGATGG 485
DB 319 GACATCTATGCAAAATATGCAAAAGAGATTAATGCTGG 359

RESULT 15

LOCUS AF060703 1557 bp mRNA PRI 31-DEC-1999
DEFINITION Homo sapiens DNAJ homolog mRNA, complete cds.
ACCESSION AF060703
VERSION AF060703.1 GI:6648622

KEYWORDS
SOURCE human.
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
AUTHORS Primates; Catarrhini; Homidae; Homo.
TITLE Zhang, J.S., Nelson, M., Wang, L. and Smith, D.I.

JOURNAL Submitted (21-APR-1998) Pathology and Lab. Medicine, Mayo Clinic,
200 SW 1st St., Rochester, MN 55905, USA

FEATURES
source 1..1557
/organism="Homo sapiens"
/db_xref="taxon:9606"

/chromosome="11"
/cell_type="pancreatic cancer cells"
/map="11q"
146..871
/note="HHDJ1"
/product="DNAJ homolog"
/protein_id="AA02157.1"

BASE COUNT 472 a 275 c 357 g 452 t 1 others
ORIGIN

Query Match 6.6%; Score 91.4; DB 40; Length 1557;
Best Local Similarity 66.5%; Pred. No. 3.3e-11;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATAAGATCTTGGGGTGCCTCGAAGTCCCTCTATAAGATATTAAAGGCC 327
DB 152 GATTACTATGAAGTTCTAGCGCGTGCAAGACATGCCACCGAGATATTAAAGGCA 211
QY 328 TATAGAACTAGCCCTGACGCTTCATCCGACCGGAACCTG--ATGATCCACAAGCC 384
DB 212 TATCGGAACTGGCAGCTGCAAGTGCATCAATTAATCTGAGATTAAGAAAGCA 271
QY 385 CAGAGAAATTCAGATCTGGGTGCTCTATGAGGTTCTGATAGTGAAGAACGG 444
DB 272 GAGAGAAATTCAGAGCAAGTAGCGAGGCAATGAGTGTGCGATGCTTAAGAACGG 331
QY 445 AAACAGTACGATCTTATGCTGAAGAGGATTTAAAGATGG 485
DB 332 GACATCTATGCAAAATATGCAAAAGAGATTAATGCTGG 372

Search completed: June 30, 2000, 12:46:48
Job time: 10107 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 12:12:13 ; Search time 929.99 Seconds
(without alignments)
5796.607 Million cell updates/sec

Title: US-09-501-714-4
Perfect score: 1330
Sequence: 1 CGNAGGAGAGAAAGAAAG.....TGACCGCACGGGTGCGGGG 1330

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 08
Listing first 45 summaries

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *

45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
59: gb_est33: *
60: gb_est34: *
61: gb_est35: *
62: gb_est36: *
63: gb_est37: *
64: gb_est38: *
65: em_est27: *
66: em_est28: *
67: em_est29: *
68: em_est30: *
69: gb_est39: *
70: gb_est40: *
71: gb_est41: *
72: gb_est42: *
73: gb_est43: *
74: gb_est44: *
75: em_est31: *
76: em_est32: *
77: em_est33: *
78: em_est34: *
79: gb_est45: *
80: gb_est46: *
81: gb_est47: *
82: gb_est48: *
83: gb_est49: *
84: gb_est50: *
85: gb_est51: *
86: em_est35: *
87: em_est36: *
88: em_est37: *
89: em_est38: *
90: gb_est52: *
91: gb_est53: *
92: gb_est54: *
93: gb_est55: *
94: gb_est56: *
95: em_est39: *
96: em_est40: *
97: em_est41: *
98: em_est42: *
99: em_est43: *
100: em_est44: *
101: em_est45: *
102: gb_est57: *
103: gb_est58: *
104: em_est46: *
105: gb_est59: *
106: gb_est60: *
107: gb_est61: *
108: gb_est62: *
109: gb_est63: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

Db	311	TGTGGAGTCGTAAGAAAGGAGACATCTATGCAAAATATATGCGAAAGAGATTAAATGGTG	370
QY	383	GNNGNGNGGTGGAAGTCATTTTGACAGCTCCATTTGAAATTTGGCTTCACATTCGTAAC	442
Db	371	GAGAGAGAGGTGGGAAGTCATTTTGACAGCTCCATTTGAAATTTGGCTTCACATTCGTAAC	430
QY	443	CAGATGATGTCCTTACAGGAAATTTTGGTGGGAAGGAGCCCATTTTCATTTGACATTCCTTG	502
Db	431	CAGATGATGTCCTTACAGGAAATTTTGGTGGGAAGGAGCCCATTTTCATTTGACATTCCTTG	490
QY	503	AAGACCCCTTTTGAGAGACTCTTTGGGAATCG-AAAGGGTCCCGAGGAAGCAAGACCGGA	561
Db	491	AAGACCCCTTTTGAGAGACTCTTTGGGAATCGAAAGGGTCCCGAGGAAGCAAGACCGGA	550
QY	562	GGGACGGGGGTCTTTTCTCTGCGTTCAGTGA-TTTCGCTTTTGGAGTGGATTTTC	620
Db	551	GGGACGGGGGTCTTTTCTCTGCGTTCAGTGGATTTTCCGCTTTTGGAGTGGATTTCT	609
QY	621	TTCTTTTATACAGAGATTACTTCTATTTGGGTGCTACGTACAGGGGGCC	670
Db	610	TTCTTTTATACAGAGATTACTTCTATTTGGGTGCTACGTACAGGGGGCC	659
RESULT	3		
AA020916			
LOCUS	AA020916	646 bp	mRNA
DEFINITION	z664h08.1	Soares retina N2b4HR	Human sapiens CDNA clone
ACCESSION	IMAGE:363807	5' similar to PIR:S23509	S23509
VERSION	AA020916	1	GI:1484660
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Mammalia;		
AUTHORS	Eutheria: Primates: Catarrhini: Hominiidae: Homo.		
	1 (bases 1 to 646)		
	Haller, L., Lennon, G., Becker, M., Donaldo, M.F., Chiapelli, B.,		
	Chisoe, S., Dietrich, N., Dubuque, T., Favello, A., Gish, W.,		
	Hawkins, E., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,		
	Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,		
	Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,		
	Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.		
	and Marini, M.		
TITLE	Generation and analysis of 280,000 human expressed sequence tags		
JOURNAL	Genome Res. 6 (9), 807-828 (1996)		
MEDLINE	97044478		
COMMENT	On May 8, 1995 this sequence version replaced gi:799404.		
	Contact: Wilson RK		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@wustl.edu		
	This clone is available royalty-free through LINT; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert length: 1524 Std Error: 0.00		
	Seq primer: -28M13 rev2 from Amersham		
	High quality sequence stop: 428.		
FEATURES	Location/Qualifiers		
source	1..646		
	/organism="Homo sapiens"		
	/db_xref="GDB:1280510"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:363807"		
	/clone_1p="Soares retina N2b4HR"		
	/sex="male"		
	/tissue_type="retina"		
	/dev_stage="55 year old"		
	/lab_host="DH10B (ampicillin resistant)"		
	/note="Organ: eye; Vector: p773D (Pharmacia) with a		
	modified polylinker; Site:1 Not I; Site:2 Eco RI; 1st		
	strand cDNA was primed with a Not I - oligo(dT) primer [5'		

BASE COUNT	168 a	141 c	177 g	156 t	4 others
ORIGIN					

39.7%; Score 527.4; DB 27; Length 646;
 Mismatch 94.6%; Pred. No. 1.4e-126;
 Conservative 0; Mismatches 28; Indels 5; Gaps 4;

CCGCGCAGACACAGCGGCACACANTTCCTGGNGCNTGAGGAGATTCGGGCCGT 85
||||| ||||| ||||| ||||| ||||| |||||
CCGCGCACACACAGCGGCACATTCTTGGAGCTGTGAGGAGACNNNGGCCGT 63

CCCCCTGCTCCCGCACC66-CGTTCTGTCTCGACCAATTCACA
CCCCCTGCTCCCGCACC66-CGTTCTGTCTCGACCAATTCACA
CCCCCTGCTCCCGCACC66-CGTTCTGTCTCGACCAATTCACA
CCCCCTGCTCCCGCACC66-CGTTCTGTCTCGACCAATTCACA

CATGGTGGATTACTATGAAGTCTAGGCGTGCAGAGACATGCCTACCGC 182
 CATAGTGGATACATAGAGTCTAGGCGTGCAGAGACATGCCTACCGC 205

AAAGGCATATCGGAACCTGGCACTGAAGTGGCATCAAGATATAAAATCCTG 242

CGAAGCAGACGAAATTCAMGCAGTAGCGAGCATATGAATGCTGT 302

CGGAGGGACATCTATGCACAATATGTGCCAAGAAGCATTAATGCTGGAC 362

[illegible]

GACCTTCCTTTGGG - AATCGAAGGGGTCCCCGAGCAACCAGACAAAAAGC 563
 CAGGGAATTTTTTTGGTGGAAGGA - CCATTTCATTGACTCTTTGAAG 481

TTTTTCTCTCGCGGTTCAGTGGATTTCGCTCTTTTGGAAAGTGGAATTTTCTCTC 623
GGACACTTCCCTTGGGGAAATCGAANGGGGTCCCCGAGGAACACAAAGCCGAAG 541

GGATT 638
TTTTTCTCGGCGTTCACTGCAATTCGGTCTTTGGAAITGGA-TTTCNTC 600

GARY 615

553 bp mRNA EST 11-AUG-1997
1 Soares_testis_NHT Homo sapiens cDNA clone IMAGE:757147

1 GI:2229426

ORGANISM	Human sapiens
REFERENCE	Eukaryota: Metazoa: Chordata: Craniata: Vertebrata; Mammalia: Eutheria: Primates: Catarrhini: Hominoidei: Homo.
AUTHORS	1 (bases 1 to 553) Hillier, J., Allen, M., Bowles, L., Dubuque, T., Gelsel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, G., Marra, M., Martin, J., Moore, B., Schellenberg, R., Steptoe, M., Tan, F., Theisig, B., White, Y., Wyllie, T., Waterston, R. and Wilson, R.
TITLE	WashU-Merck EST Project 1997
JOURNAL	Unpublished (1997)
COMMENT	Contact: Wilson RK Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available royalty-free through INM; contact the INM Consortium (info@inm.jku.at) for further information. Seq primer: -28m13 rev2 Err from Amersham High quality sequence stop: 454. Location/Qualifiers
FEATURES	

```

/organism="Homo sapiens"
/db_xref="GDB:597775"
/db_xref="taxon:9606"

```

```
/clone_lib="Soares_testis_NHT"  
/sex="male"  
/lab_host="DH10B"  
/host_vector="EM733D_pos (Pharmacia) with a 3.0 kb"
```

polylinker: Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5',

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTZ19 vector library (1615ACCAATCGAGCGGACCGCCGCCAATTTTTTTTTT 3').

BASE COUNT	ORIGIN
150 a	113 c 151 g 139 t

Query Match	39.4%;	Score 523.8;	DB 34;	Length 553;
Best Local Similarity	98.6%;	Pred. No. 1.2e-12;		

[illegible]

124 TTCCTGGAGCCATTCACACATCTCGTAAACATGCTGATTTACTATGAAGTTCAGGC 183

QY 184 GTGCAAGACATGCCCTCACCCGAGATATTAAAAAGCATATCGAACTGGCATTGAAG 243
 |||||
 120 GTGCAGAGACATGCCCTCACCCGAGATATTAAAAAGCATATCGAACTGGCATTGAAG 170

QY 244 TGGCATCCAGTAAATAATCTCTGAGAAATAAGAGAAGCAGAGAGAAATTCAGCAAGTA 303
|||||
180 TGGCATCCAGTAAATAATCTCTGAGAAATAAGAGAAGCAGAGAGAAATTCAGCAAGTA 320
|||||

QY 304 GCGAGGCATATGAAGTGGCTGTCGGATGCTAAGAAACGGGCATCTATGACAAATATGGC 363
||||| |||||||||
db 240 GCGGG-GCATATGAAGTGCTCGGATGCTAAGAAGCGGACATCTATTGCAAAATATGGC 208

[illegible]

|||||
Db 359 GGCCTTCATTCGCGTAACCCAGATGATCTCTTCAGAGGAATTTTGGGGAAGGACCA 418
OY 484 TTTTCATTGACTCTCTTTGAAGACCCCTTTGAGGACTCTTTGGGAATCGAAGGGCTCC 543
Db 419 TTTTCATTGACTCTCTTTGAAGACCCCTTTGAGGACTCTTTGGGAATCGAAGGGCTCC 478
OY 544 CGAGAACCAAGACCGGAGGAGCGGGTCTTTCTGCTGCTTGTAGTGGATTTCCTCT 603
Db 479 CGAGAACCAAGACCGGAGGAGCGGGTCTTTCTGCTGCTTGTAGTGGATTTCCTCT 538
OY 604 TTTGGAAGTGGATT 618
Db 539 TTTGGAAGTGGATT 553
RESULT 5
AA993699/c
LOCUS
DEFINITION AA993699 529 bp mRNA EST 27-AUG-1998
OC97C10.s1 Soares_total_fetus.ND2HF8_9w Homo sapiens cDNA clone
IMAGE:1624722 3' similar to TR:035723 035723 TESTIS SPECIFIC
DNAJ-HOMOLOG.; contains element MSRI MSRI repetitive element ;,
mRNA sequence.
ACCESSION AA993699 GI:3180244
VERSION AA993699.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 529)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 700 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 457.
FEATURES
Source
Location/Qualifiers
1..529
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1624722"
/clone_lib="Soares_total_fetus.ND2HF8_9w"
/dev_stage="8-9 weeks"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from mRNA obtained from pooled 8-9 week
(total) fetus material with a Not I - oligo(dT) primer [5'
TGTACCAATCGAAGTGGAGCGGCGGCTTAATTTTCTTTTCTTTT 3']
Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 145 a 143 c 89 g 152 t
ORIGIN

Query Match 38.1%; Score 506.6; DB 40; Length 529;
Best Local Similarity 97.7%; Pred. No.3.4e-121;
Matches 512; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

OY 157 ATGTGATTAATGAAATCTAGCGCTGCAGACATGCTCACCGAGGATTTAA 216
|||||
Db 529 ATGTGATTAATGAAATCTAGCGCTGCAGACATGCTCACCGAGGATTTAA 470

OY 217 AAGCATATCGGAACCTGGACCTAGTGGCATCCAGATTAATAATCTGAGATAAAGA 276
|||||
Db 469 AAGCATATCGGAACCTGGACCTAGTGGCATCCAGATTAATAATCTGAGATAAAGA 410
OY 277 GAAGCAGAGAGAAATTCAGCAATAGCGAGGATATGAGTCTGCGATCTTAAG 336
|||||
Db 409 GAAGCAGAGAGAAATTCAGCAATAGCGAGGATATGAGTCTGCGATCTTAAG 350
OY 337 AAACGGACATCTATGACAAATATGCGCAAGAAGATTAAATGTGNGNGNGGTGA 396
Db 349 AAACGGACATCTATGACAAATATGCGCAAGAAGATTAAATGTGNGNGGTGA 290
OY 397 AGTCATTTTGGACAGTCATTTTGAATTTGGCTTCACATTCGCCGACAGATATCTTC 456
Db 289 AGTCATTTTGGACAGTCATTTTGAATTTGGCTTCACATTCGCCGACAGATATCTTC 230
OY 457 AAGCAATTTTGTGGGAAGGACCATTTTCACTTGTGATTCCTTTGAGACCTTTGAG 516
Db 229 AAGCAATTTTGTGGGAAGGACCATTTTCACTTGTGATTCCTTTGAGACCTTTGAG 170
OY 517 GACTTCTTTGGGAATCGAAGGGGTCGCCGAGAGACAGAACCGAGGAGTCTGTT 576
Db 169 GACTTCTTTGGGAATCGAATTTTTCGCCGAGAGACAGAACCGAGGAGTCTGTT 110
OY 577 TTCTCTGCGTTAGTGGATTTCCTGCTTTTGGAAATGATTTTCTTTTGTATCAGA 636
Db 109 TTCTCTGCGTTAGTGGATTTCCTGCTTTTGGAAATGATTTTCTTTTGTATCAGA 50
OY 637 TTTACTCATTTTGGGTCACCTAGTGCAGGGGGGCTCCTTCATT 680
Db 49 TTTACTCATTTTGGGTCACCTAGTGCAGGGGGGCTCCTTCATT 6

RESULT 6
AI028519/c
LOCUS
DEFINITION AI028519 535 bp mRNA EST 28-AUG-1998
OM44b09.x1 Soares_parathyroid_tumor.NbHRA Homo sapiens cDNA clone
IMAGE:1649657 3' similar to TR:035723 035723 TESTIS SPECIFIC
DNAJ-HOMOLOG.; mRNA sequence.
ACCESSION AI028519
VERSION AI028519.1 GI:3245828
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 535)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
www-bio.llnl.gov/dbtrp/image/image.html
Insert Length: 668 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 450.
FEATURES
Source
Location/Qualifiers
1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1649657"
/clone_lib="Soares_parathyroid_tumor.NbHRA"
/tissue_type="parathyroid tumor"

BASE COUNT	179 a	147 c	175 g	152 t	8 others
ORIGIN	TGTACCAATCTGAAGTGGAGGCGGCGGACCAATTTTTTTTTTTTTTTTTTTTTTTT T-3'), double-stranded cDNA was size selected, ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7n3 vector (Pharmacia). Library went through one round of normalization to a Cot = 5. Library constructed by Bento Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid adenomas was kindly provided by Dr. Stephen Marx, National Institute of Diabetes and Digestive and Kidney Diseases, NIH."				
Query Match	36.7%; Score 488; DB 26; Length 661;				
Best Local Similarity	93.6%; Pred. No. 2,5e-116;				
Matches	559; Conservative 0; Mismatches 31; Indels 7; Gaps 5;				
OY	23	GCCGAGAGAGCCGCCGACACACAGCGNCACANTCCTGG-NGCTNTAGAGAGATTCGGG	81		
Db	14	GCGGAGAGAGCGCGCGCCACACACAGCCACAGNCCTCGAGNNNTGTAGGAGATTCGGG	73		
OY	82	CCGTGACCCCTGCTCCCTCCGCTCCCGGACACCGCGCTCTCTTCCCTGGAGCCCAATTCGA	141		
Db	74	CCGTGACCCCTGCTCCCTCCGCTCCCGGACACCGCGCTCTCTTCCCTGGAGCCCAATTCGA	133		
OY	142	ACAATCTGTAAACATGATGGTATGATGAAATTCATAGGCGTGACAGACATCCCTCA	201		
Db	134	ACAATCTGTAAACATGATGGTATGATGAAATTCATAGGCGTGACAGACATCCCTCA	193		
OY	202	CCCGAGGATTTTAAAAAGGCATATCGGAACTGGCAGCTGAAAGTGGCATTCAGATTTAAAT	261		
Db	194	CCCGAGGATTTTAAAAAGGCATATCGGAACTGGCAGCTGAAAGTGGCATTCAGATTTAAAT	253		
OY	262	CCTGAGATTTAAAGAAAGAGAGAGAAATTCAGACAGATGAGGAGGCATATGAGTG	321		
Db	254	CCTGAGATTTAAAGAAAGAGAGAGAAATTCAGACAGATGAGGAGGCATATGAGTG	313		
OY	322	CTGTGCGATGCTTAAAGAAAGCGGACATCTATGACAAATATGCGAAAGAGATTTAAATG	381		
Db	314	CTGTGCGATGCTTAAAGAAAGCGGACATCTATGACAAATATGCGAAAGAGATTTAAATG	373		
OY	362	GGNGGNGNGGTGGAAGTCATTTTGACAGTCCATTTGAATTTGGCTTCACATTTCCGTAA	441		
Db	374	GGAGGAGAGAGGTGGAAGTCATTTTGACAGTCCATTTGAATTTGGCTTCACATTTCCGTAA	433		
OY	442	CCAGATGATGCTTGTAGGGAATTTTGGTGGAAAGGACCAATTTTCATTTGACATCTCTT	501		
Db	434	CCAGATGATGCTTGTAGGGAATTTTGGTGGAAAGGACCAATTTTCATTTGACATCTCTT	493		
OY	502	GAAAGACCTTTTGAAGAGCATCTTTGGAGATCAAGAGGGTCCCGGAGGAAGCAGAACCGGA	561		
Db	494	GAAAGACCTTTTGAAG-ACCTCTTTGGAGATCAAGAGGGT-CCCGGAGGAAGCAGAACCGGA	550		
OY	562	GGAGACGGGGTCTTTTCTCTGCGTTCAAGTGGATTTCCGCTTTTGGAGATGATTT	618		
Db	551	AGGACGGGG-CGNTTTTCCCTGCTTCAGTGGATTT-CCGCTTTTGGAGATGATTT	604		
RESULT	9				
AL036026	531 bp mRNA EST 27-SEP-1999				
LOCUS	AL036026				
DEFINITION	DKR2P564P0422.r1.564 (synonym: h6fbr2) Homo sapiens cDNA clone				
ACCESSION	AL036026				
VERSION	AL036026.1 GI:5405652				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.				
	Wambutt, R., Haubner, D., Mewes, H.W., Gassenhuber, J. and Wiemann, S.				

	QY	254	ATGAAAACTCGAGATTAAGAAGACAGAGAAAATTCAAGCAAGTAGCGGAGGCAT	313
	Dd	181	ATAAAATCCTGGAGAAATTAAGAGAGACAGAGAAAATTCAAGCAAGTAGCGGAGGCAT	240
	QY	314	ATGAAGTCGTGTGGATGCTTAAGAAAACGGGACACTTATACAAATATGGCAAAAGAGAT	373
	Dd	241	ATGAAGTCGTGTGGATGCTTAAGAAAACGGGACACTTATACAAATATGGCAAAAGAGAT	300
	QY	374	TAAATGTGNGNGNGNGTGGAAGTCATTTTGACAGTCATTTGATTAATTTGGCTTCACAT	433
	Dd	301	TAAATGTGAGAGAGAGAGGGGTGAAGTCATTTTGACAGTCATTTGATTAATTTGGCTTCACAT	360
	QY	434	TCCGTAACCCAGATGATGTCCTTCAGGGAAATTTTTGGTGAAGGAGACCAT-TTTTCATTT	492
	Dd	361	TCCGTAACCCAGATGATGTCCTTCAGGGAAATTTTTGGTGAAGGAGACCATTTTTCATTT	420
	QY	493	GACTCTTTTAAGAACCCTTTTGAGGACTCTTTGGGAATCGAAGGGGTCCCAGAGAAC	552
	Dd	421	GACTCTTTTAAGAACCCTTTTGAGGACTCTTTTAAGAAAGGATCCCCAGAGAAC	480
	QY	553	AGAGCCGAGGAGACGGGGGTGTTTCTGTCGCCGTGAGTGAGATTTCCGCTTT	605
	Dd	481	TAAAGCCAGGAGACGGGGGTGTTTCTGTCGCCGTGAGATTTCCGCTTT	533
	RESULT	12		
	LOCUS	AU080965	804 bp mRNA EST	20-OCT-1999
	DEFINITION	AU080965 Sugano mouse brain mncb Mus musculus cDNA clone MNCB-6291		
	ACCESSION	AU080965		
	VERSION	AU080965.1	GI:6085719	
	KEYWORDS	EST.		
	SOURCE	house mouse.		
	ORGANISM	Mus musculus		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
	AUTHORS	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
	JOURNAL	1 (bases 1 to 804)		
	COMMENT	Hashimoto, K., Kusuda, J., Toyoda, A., Tanuma, R., Ito, A., Hirata, M., Suzuki, Y., Saeki, M. and Sugano, S.		
		Isolation of full-length cDNA clones from a mouse brain cDNA library made by oligo-capping method unpublished (1999)		
		On Jun 5, 1998 this sequence version replaced gi:3189734.		
		Contact: Katsuyuki Hashimoto		
		National Institute of Infectious Diseases		
		23-1, Toyama 1-chome, Shinjuku-ku, Tokyo 162-8640, Japan		
		Email: khashie@nih.go.jp		
		URL: http://www.nih.go.jp/yoken/gsnbank/.location/Qualifiers		
	FEATURES			
	source			
		1..804		
		/organism="Mus musculus"		
		/strain="C57BL"		
		/db_xref="taxon:10090"		
		/clone="MNCB-6291"		
		/clone_lib="Sugano mouse brain mncb"		
		/sex="female"		
		/dev_stage="adult"		
		/lab_host="TOP10"		
		/note="Organ: brain; Vector: pME18S-FL3; 1st strand cDNA was primed with an oligo(dT) primer [ATGCGCGCTTTTCTTTTCTTTT] double-stranded cDNA was ligated to a DraIII adaptor [TTTTCGCTTACGG], digested and cloned into distinct DraIII sites of the pME18S-FL3. XhoI sites just outside the DraIII sites can be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5 kb. Library was constructed by Sugano et al. (University of Tokyo, Institute of Medical Science). Custom primer for sequencing: 5' end primer [CTTCTGCTTAAAGTCGCG]		
	BASE COUNT	226 a	161 c	205 g
		199 t	13 others	

ORIGIN

Query Match 35.7%; Score 474.4; DB 69; Length 804;
Best Local Similarity 79.4%; Pred. No. 9.5e-113;
Matches 614; Conservative 0; Mismatches 145; Indels 14; Gaps 5;

QY 4 AGGAGAGNAAAGGAAAGGCCCGCAGAGAGCCCGCAGACCAACGACGNCACANTCCGCGN 63
DB 17 AAGAGAGAAAGGAAAGGCCCGCAGAGAGAGCCCGCAGACCAACGACGAGCTGT 76
QY 64 GCTNTGAGAGATTCGGGGCGTCAACCTGCTCCCTGCTTCCCGCAGCCGCGCTTCT 123
DB 77 GCCACAGCCCGCAGAGAGAGCCGCGCTGACTTCCTCCCTGCTGCGCGCGCG-----T 129
QY 124 TTCCTGGAGCCATTCACAACTCTGTAAACATGCTGATTAATGAAAGTTCTAGGC 183
DB 130 GTCTCGGATTTATTCACAGTCAGTTAAACATGCTGATTAATGAAAGTTCTAGGC 189
QY 184 GTGACAGACATGCTTACCGCAGAGATTTAAAGGATATGGAACATGCGAACGCTGAG 243
DB 190 GTGACAGACATGCTTACCGCAGAGATTTAAAGGATATGGAACATGCGAACGCTTAA 249
QY 244 TGGCATTCAGATTAATAATCCTGAGATTAAGAGAGAGAGAGAGAGAGAGAGAGTA 303
DB 250 TGGCAGCCCGAGTA 309
QY 304 GCGAGGACATATGATGCTGCTGCTGATGCTTAAGAACGAGACATCTATGACAAATATGCG 363
DB 310 GCTGAGGACATATGATGCTGCTGATGCTTAAGAACGAGACATCTATGACAAATATGCG 369
QY 364 AAGAGAGATTAATA--TGGTGGNGNGNGGTGAGAGTCAATTTGACAGTCCATTGAA 420
DB 370 AAGAGAGATTAATAATGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
QY 421 TTTGGCTTCAATTCCTTAACCCAGATGATGCTTCAAGGAAATTTTGTGAGAGGAC 480
DB 430 TTTGGCTTCAATTCCTTAACCCAGATGATGCTTCAAGGAAATTTTGTGAGAGGAC 489
QY 481 CCATTTTCAATTTGATCTTTGAAGACCTTTGAGAGCTTTTGGAAATCGAAGGGGT 540
DB 490 CCATTTTCAATTTGATCTTTGAAGACCTTTGATGATGATCTTTTGGAAATCGAAGGGGT 549
QY 541 CCCGAG 600
DB 550 CCCGAG 609
QY 601 TCTTTTGAAGTGAATTTCTTTTGAATGACAGATTTACTTCAATTTGGGTCACTAGCT 660
DB 610 TCTTTTGAAGTGAATTTCTTTTGAATGACAGATTTACTTCAATTTGGGTCACTAGCT 669
QY 661 CA-CGGGGGCTTCAATTCCTTCAAGTCAATTTGGGTGAGTGGAGAGAGAGAGAG 719
DB 670 CATTGGGGCTTCAATTCCTTCAAGTCAATTTGGGTGAGTGGAGAGAGAGAGAGAG 728
QY 720 CAATTCGATATCAACTCAATAAATGTTATGAGAGAGAGAGAGAGAGAGAGAGAG 772
DB 729 NAATNTATTTTNA--CTTCACTNMTATGATGAGAGAGAGAGAGAGAGAGAGAGAG 779

RESULT 13
AL036764 535 bp mRNA EST 27-SEP-1999
LOCUS DKEF2564J1863.1 564 (synonym: hibr2) Homo sapiens cDNA clone
DEFINITION DKEF2564J1863.1, mRNA sequence.
ACCESSION AL036764
VERSION AL036764.3 GI:5927903
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 535)

AUTHORS Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann, S.
TITLE EST (Duesterhoeft, et al.)
JOURNAL Unpublished (1999)
COMMENT On Jul 7, 1999 this sequence version replaced gi:5866299.
Contact: Duesterhoeft A

AM Kiofespitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ), Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Clagen within the cDNA sequencing consortium of the
German Genome Project.
No si sequence available.
This clone is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
Location/Qualifiers
1. 535

FEATURES
source
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZP564J1863"
/clone_1ib="564 (synonym: hibr2)"
/tissue_type="brain"
/dev_stage="fetal"
/lab_host="X1-2blue"
/note="Vector: PAMPI; Site:1: NotI; Site:2: SalI"

BASE COUNT 148 a 117 c 145 g 117 t 8 others
ORIGIN

Query Match 34.9%; Score 464; DB 64; Length 535;
Best Local Similarity 94.4%; Pred. No. 4e-110;
Matches 493; Conservative 0; Mismatches 26; Indels 3; Gaps 2;

QY 23 GCCGAG 82
DB 16 GCGGAG 75
QY 83 CGTACCCCTGCTCCCTGCTTCCCGCAGCGGCGCTTCTTCTGAGAGAGAGAGAGAG 142
DB 76 CGTACCCCTGCTCCCTGCTTCCCGCAGCGGCGCTTCTGCTGAGAGAGAGAGAGAG 135
QY 143 CAATCTGTAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 202
DB 136 CANTCTGTAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 195
QY 203 CCGAGATATTAAG 262
DB 196 CCGAGG--ATTAAAG 253
QY 263 CTGAGATTAAG 322
DB 254 CTGAGATTAAG 313
QY 323 TGTGATGCTTAAG 382
DB 314 TGTGATGCTTAAG 373
QY 383 GNGGNGNGGTGAG 442
DB 374 GAG 433
QY 443 CAGATGATGCTTCAAG 502
DB 434 CAGATGATGCTTCAAG 493
QY 503 AA-GACCTTTTGAAG 543
DB 494 AAG 535

RESULT 14

performed to exclude fragments <1.5kb. Library
constructed by Dr. Sumio Sugano (University of Tokyo
Institute of Medical Science). Custom primers for
sequencing: 5' end primer CTCTGCTCTTAAAGCTGGC and 3' end
primer CGACCTGCAGCTCGACACA."

BASE COUNT 172 a 148 c 175 g 160 t
ORIGIN

Query Match 34.0%; Score 452; DB 41; Length 655;
Best Local Similarity 82.8%; Pred. No. 5.8e-107;
Matches 552; Conservative 0; Mismatches 99; Indels 16; Gaps 3;

```
QY 12 AAAGGAAAGNCGCCGAGAGCCGCCGCMACACAGCGNCACANTCCTGGNGCTNTGAG 71
Db 2 AAGCGCAGAGAGAGCCGCCGCCGCAACCGCACAGGCGAGCTGTGCGACAGCCGCGAG 61
QY 72 GAGATTGGGCGGCTCACCTCCCTCCCTGCTCCGCGCACCGGCGGCTTCTCTGG 131
Db 62 GAGA--CGGGCGGCTGACTCTCCCTCGCTGGCGCC-----GTGTCCTCGG 108
QY 132 ACCCATTCCAACATCTCGTAAACATGTGATTAAGTCTAGGCGTGCAGAG 191
Db 109 ATTATTCCAAACAGTCACTTAAACATGTGATTAAGTCTAGGCGTGCAGAG 168
QY 192 ACATGCTCAACCCGAGAGATATTTAAAGCATATCGAAACTGGCACTGAAGTGCATCC 251
Db 169 ACATGCTCAACCTGAGAGACATTTAAAGGGGTATCGAAACAGGCACTTAATGGCACCC 228
QY 252 AGATAAATTCCTGAGAAATTAAGAAAGCAGAGAAATTCAGCAAGTAGCGAGGC 311
Db 229 GGACAAAATTCCTGAAATTAAGAAAGCAGAGAAATTCAGCAAGTAGCTGAGGC 288
QY 312 ATATGAAGTCTGTGAGATCTAAGAAAGGACATCTATGACAAATATGCAAGAAGG 371
Db 289 ATATGAAGTGTATGAGATCTAAGAAAGGACATCTAGCACAATATGCAAGAAGG 348
QY 372 ATTAAA---TGTGNGNGNGNGGTGGAAGTCTTTGACAGTCCATTTGAATTTGGCTT 428
Db 349 ATTAAATGTGTGAGAGAGAGGTGGAATTTGACAGTCCATTTGAGTTGGCTT 408
QY 429 CACATTCGTAACCCAGATATGTCTCAGGAAATTTTGTGGAAGGACCATTTTC 488
Db 409 CACATTCGGAACCCAGATATGTCTCAGGAAATTTTGTGGAAGGACCATTTTC 468
QY 489 ATTGACTCTTTGAAGACCTTTTGAGACTTCTTGGAATGGAAGGGGTCCCGAGG 548
Db 469 ATTGACTTCTTTGAAGACCCATTTGATGACTTTTGGAAACGGAAGGGGTCCCGAGG 528
QY 549 AAGCAGAAGCCGAGAGCGGGGTCTCTGCGTCAAGTGAATTCGGTCTTTGG 608
Db 529 AAATAGAAAGCCGAGAGTCCGGCTCATTTTCTACCTCAGTGATTTCCCTTTTGG 588
QY 609 AAGTGAATTTCTTCTTTTATACAGATTTACTTCAATTTGGTCACTAGGTACGGGG 668
Db 589 AAGTGAATTTCTTCTTTTATACAGGCTTCACTCATTTGGGTCACTAGGTATGGGG 648
QY 669 CTTCACT 675
Db 649 TCTCACT 655
```

Search completed: June 30, 2000, 12:12:25
Job time: 8076 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 09:57:49 ; Search time 929.99 Seconds

(without alignments)
5997.092 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376

Sequence: 1 TCTCACC GGACTCGGACT.....TTTATTTCATATGCAACT 1376

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 202611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST: *
1: em_est1: *
2: em_est2: *
3: em_est3: *
4: em_est4: *
5: em_est5: *
6: em_est6: *
7: em_est7: *
8: em_est8: *
9: em_est9: *
10: em_est10: *
11: em_est11: *
12: em_est12: *
13: em_est13: *
14: em_est14: *
15: em_est15: *
16: em_est16: *
17: em_est17: *
18: em_est18: *
19: em_est19: *
20: gb_est1: *
21: gb_est2: *
22: gb_est3: *
23: gb_est4: *
24: gb_est5: *
25: gb_est6: *
26: gb_est7: *
27: gb_est8: *
28: gb_est9: *
29: gb_est10: *
30: gb_est11: *
31: gb_est12: *
32: gb_est13: *
33: gb_est14: *
34: gb_est15: *
35: gb_est16: *
36: gb_est17: *
37: gb_est18: *
38: gb_est19: *
39: gb_est20: *
40: gb_est21: *
41: gb_est22: *
42: gb_est23: *
43: gb_est24: *
44: gb_est25: *

45: gb_est26: *
46: gb_est27: *
47: gb_est28: *
48: gb_est29: *
49: gb_est30: *
50: gb_est31: *
51: gb_est32: *
52: em_est20: *
53: em_est21: *
54: em_est22: *
55: em_est23: *
56: em_est24: *
57: em_est25: *
58: em_est26: *
59: gb_est33: *
60: gb_est34: *
61: gb_est35: *
62: gb_est36: *
63: gb_est37: *
64: gb_est38: *
65: em_est27: *
66: em_est28: *
67: em_est29: *
68: em_est30: *
69: gb_est39: *
70: gb_est40: *
71: gb_est41: *
72: gb_est42: *
73: gb_est43: *
74: gb_est44: *
75: em_est31: *
76: em_est32: *
77: em_est33: *
78: em_est34: *
79: gb_est45: *
80: gb_est46: *
81: gb_est47: *
82: gb_gss1: *
83: gb_gss2: *
84: gb_gss3: *
85: gb_gss4: *
86: gb_gss1: *
87: em_gss2: *
88: em_gss3: *
89: em_gss4: *
90: gb_gss5: *
91: gb_gss6: *
92: gb_gss7: *
93: gb_gss8: *
94: gb_gss9: *
95: em_gss5: *
96: em_gss6: *
97: em_gss7: *
98: em_gss8: *
99: em_gss9: *
100: em_gss10: *
101: em_gss11: *
102: gb_gss10: *
103: gb_gss11: *
104: em_gss12: *
105: gb_gss12: *
106: gb_gss13: *
107: gb_gss14: *
108: gb_gss15: *
109: gb_gss16: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result

Query

OY	1069	GATTAGATCACACGCGCCAGGAGCAAAANTANPTGGGAAGAAGGGGCTGCCAACACTT	1122
Db	541	CATTAGATCACCAGGCCANGAGCCGAAAGCTATGTGAAGAAAAGGGGAAGGCTCCCCAAGTTT	600
OY	1123	GACAACAACAATATCATAGGCGCTCTTGATTAATCACT- TTGTGTGTGATTTTTCCAAAAA	1181
Db	601	GACACAACAATAATATCAAGGGCTCTTNNGATTAATCACTATTGTGATGTGGATTTTCAAAAA	660
OY	1182	ACAGTTAACAGAGGAGGAGGAGGAAAGGTATCAAACAGCTACTGAAACAAGGCTCAGTGCA	1241
Db	661	ACCAGTTCACAGTAGGAGGAGGAGAGAAAGT-TCTAACAGCTACTGAAACAAGGCTCAGTGCA	719
OY	1242	GAAAGTATCAATATGC 1256 1	
Db	720	AAGGSTATRCATATGC 734	
RESULT	2		
LOCUS	AM027159/c		
DEFINITION	Wt2db07.x1 Soares_thymus_NHFrH Homo sapiens cDNA clone IMAGE:2512981 3' similar to wp:T15H9.1 CE01664 HYPOTHETICAL 39.9 KB PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ; , mRNA sequence.		
ACCESSION	AM027159		
VERSION	AM027159.1 GI:5885915		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 656) NCI-CGAP http://www.ncbi.nlm.nih.gov/hncigap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index Unpublished (1997)		
JOURNAL	On Jun 5, 1998 this sequence version replaced gi:3189254. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov This clone is available royalty-free through LINTL : contact the IMAGE Consortium (info@image.llnl.gov) for further information. Seq primer: -400P from Gldco High quality sequence stop: 452.		
FEATURES	Location/Qualifiers 1..656 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="IMAGE:2512981" /clone_1lb="Soares_thymus_NHFrH" /dev_stage="fetal" /lab_host="DH10B (phage-resistant)" /note="Organ: thymus; pooled; Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA was primed with a Not I oligo(dT) primer [5' TGTTACCAATCGAATGGGAGCGGCGCCAGCAAGCTTTTTTTTTTTTTTTT 3'] TGGTACCACATCGAATGGGAGCGGCGCCAGCAAGCTTTTTTTTTTTTTTTT 3'] double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization. Library constructed by Bento Soares and M. Fatima Bonaldo.		
BASE COUNT	150 a 185 c 114 g 205 t 2 others		
ORIGIN			
Query Match	45.6%; Score 637.2; DB 64; Length 656;		
Best Local Similarity	98.0%; Pred. No. 2e-144;		
Matches 643; Conservative	0; Mismatches 12; Indels 1; Gaps 1.		
OY	643	TTTGTGGAAGTGTGTAGAACAAACCTGTGGGAAGGAGCGCTCGGCAAAAGGAAGTGC 702 	
Db	656	TTTGTGTGAGTGTGTAGAACAAACCTGTGGGAAGGAGCGCTCGGCAAAAGGAAGTGC 597	

Y	703	AAATGTGGGCAAGAGATGCGGACCCAGCGAGTGGGGCCCTGGGGCGCTTCCAAATGACCCAG	762
Y <td>703</td> <td>AAATGTGGGCAAGAGATGCGGACCCAGCGAGTGGGGCCCTGGGGCGCTTCCAAATGACCCAG</td> <td>762</td>	703	AAATGTGGGCAAGAGATGCGGACCCAGCGAGTGGGGCCCTGGGGCGCTTCCAAATGACCCAG	762
Db	596	AAATGTGGGCAAGAGATGCGGACCCAGCGAGTGGGGCCCTGGGGCGCTTCCAAATGACCCAG	537
Y	763	GAGGTGTCTCTGGAGAGATGCGCTTAATCTCAAACTAGTGAATGAAGAAGCAAGCGCTGGAA	822
Db	536	GAGGTGTCTCTGGAGAGATGCGCTTAATCTCAAACTAGTGAATGAAGAAGCAAGCGCTGGAA	477
Y	823	GTCAGAAATAGACCTGGGGGTGAGAGACGGCATGCGATACCCCTTATTTGAGAGAGGTGAG	882
Db	476	GTCAGAAATAGACCTGGGGGTGAGAGACGGCATGCGATACCCCTTATTTGAGAGAGGTGAG	417
Y	883	CCTCAGCGGATGGGGGAGCCTGGAGATTTAGGGTTCCGAAATCAAAAGTGTCAAGCACCA	942
Db	416	CCTCAGCGGATGGGGGAGCCTGGAGATTTAGGGTTCCGAAATCAAAAGTGTCAAGCACCA	357
Y	943	ATATTTGAAAGAGAGAGATGATTTTGTACACAAATGTGACAGTCTCATTTAGTTAGTCA	1002
Db	356	ATATTTGAAAGAGAGAGATGATTTTGTACACAAATGTGACAGTCTCATTTAGTTAGTCA	297
Y	1003	CTGCTTGGCTTTGAGATGATTAATCTACACTTTGATGCTCACAGGTACATA-TTTCCCG	1061
Db	296	CTGCTTGGCTTTGAGATGATTAATCTACACTTTGATGCTCACAGGTACATA-TTTCCCG	237
Y	1062	GGATTAAGTTCACCGAGCGGACGAGCAAGTATMTNGAAGAAAGGGAAGGGGCTCCCAACTT	1121
Db	236	GGATTAAGTTCACCGAGCGGACGAGCAAGTATMTNGAAGAAAGGGAAGGGGCTCCCAACTT	177
Y	1122	TGACACAACAATATCAAGGGGCTTTGTATATATCACTTTTGTATGTGATTTTCCAAAAGA	1181
Db	176	TGACACAACAATATCAAGGGGCTTTGTATATATCACTTTTGTATGTGATTTTCCAAAAGA	117
Y	1182	ACAGTTAACAGAGGAGGAGGAGAGAGATATCAACAGCTACTGAAACAAAGGTGAGTCA	1241
Db	116	ACAGTTAACAGAGGAGGAGGAGAGATATCAACAGCTACTGAAACAAAGGTGAGTCA	57
Y	1242	GAAGTTATACATGAGTCTGCAAGGATTTTGGAGAGTGAATTAATTTGCACTTTGTTT	1297
Db	56	GAAGTTATACATGAGTCTGCAAGGATTTTGGAGAGTGAATTAATTTGCACTTTGTTT	1
RESULT	3		
LOCUS	AA639658/c		
DEFINITION	ng3907.s1 NCI CGAP Co9 Homo sapiens cDNA clone IMAGE:1158972 3'		
ACCESSION	AA639658		
VERSION	AA639658.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Homiidae; Homo.		
TITLE	1 (bases 1 to 695)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),		
	Tumor Gene Index		
	Unpublished (1997)		
	On May 9, 1995 this sequence version replaced gi:802630.		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	Tissue Procurement: Ilan Kirsh, M.D., Michael R. Emmert-Buck,		
	M.D., Ph.D.		
	cDNA Library Preparation: M. Bento Soares, Ph.D.		
	cDNA Library Arrayed by: Greg Lennon, Ph.D.		
	DNA Sequencing by: Washington University Genome Sequencing Center		
	clone distribution: NCI-CGAP clone distribution information can be		
	found through the I.M.A.G.E. Consortium/ILNL at:		
	www-bio.llnl.gov/dbp/image/image.html		

Oy	808	GAGCAACGCTGGAAAGTGAATATGAGCTGGGGGAGACAGCGCATGAGTACCCCTT	867
Db	415	GACGACACCGTGGAGTGAATATGAGCTGGGGTGAAGACGGCATGAGTACCCCTT	416
Oy	868	ATTGGAGAAGGTGAGCCCTCAGCTGGATGGGAGCCCTGGAGATTTATACGGTTCGATCAAA	927
Db	415	ATTGGAGAAGGTGAGCCCTCAGCTGGATGGGAGCCCTGGAGATTTATACGGTTCGATCAAA	356
Oy	928	GTGTGCAAGCACCCAAATATTTGAAAAGGAGAGAGATGATTTGTACACAAATGTGACAGTC	987
Db	355	GTGTGCAAGCACCCAAATATTTGAAAAGGAGAGAGATGATTTGTACACAAATGTGACATC	296
Oy	988	TCATAGTGTGAGTCACTGCTGGCTTTCAGATNGATATTAATCTCACTTGGATGGTCAACAG	1047
Db	295	TCATAGTGTGAGTCACTGCTGGCTTTCAGATNGATATTAATCTCACTTGGATGGTCAACAG	236
Oy	1048	GTACATATTTCCCGGATTAAGTACACACGCGCAGAGCGAANTNTGGAAAGAAAGGGAA	1107
Db	235	GTACATATTTCCCGGATTAAGTACACACGCGCAGAGCGAANTNTGGAAAGAAAGGGAA	176
Oy	1108	GGGCTCCCCAATTTGACACACACAAATATCAAGGGCTCTTTGATTAATCACTTTTGATGTG	1167
Db	175	GGGCTCCCCAATTTGACACACACAAATATCAAGGGCTCTTTGATTAATCACTTTTGATGTG	116
Oy	1168	GATTTTCCAAAAGAACATTTAACAGAGAGAACGAGAGAGATATCAACACACTCTGATA	1227
Db	115	GATTTTCCAAAAGAACATTTAACAGAGAGAACGAGAGAGATATCAACACACTCTGATA	56
Oy	1228	CAAGGCTAGTGCAGAGAGTATACAAATGAGCTGCAGAGGATTTGAGAGTGAATA	1281
Db	55	CAAGGCTAGTGCAGAGAGTATACAAATGAGCTGCAGAGGATTTGAGAGTGAATA	2
RESULT	5		
LOCUS	AA59885/c		
DEFINITION	aa59885 498 bp mRNA EST 12-MAY-1999		
IMAGE	1091276 3' similar to WP:R15H9.1 CE01664 DNAJ ; mRNA sequence.		
ACCESSION	AA59885		
VERSION	AA59885.1		
KEYWORDS	GI:2433510		
SOURCE	EST.		
ORGANISM	human.		
REFERENCE	Homo sapiens		
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo. 1 (bases 1 to 498)		
TITLE	Jia, L., Robey, P., Young, M., Schuler, G., Powell, J., Yang, L., Lennon, G., Hillier, L., Allen, M., Bowles, L., Geisel, S., Kucba, T., Maira, M., Martin, J., Stepien, M., Tan, F., Theising, B., Bowers, Y., Wylie, T., Waterston, R., Wilson, R. and Francomano, C.		
JOURNAL	WashU-MGB/NHRI EST Project		
COMMENT	Unpublished (1997)		
	On May 9, 1995 this sequence version replaced gi:802280.		
	Contact: Wilson RK / Jia L		
	WashU-MGB/NHRI EST Project		
	Washington University School of Medicine		
	4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA		
	Tel: 314 286 1800		
	Fax: 314 286 1810		
	Email: est@watson.wustl.edu		
	This clone is available royalty-free through LINT ; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Possible reversed clone: similarity on wrong strand		
	Insert Length: 1376 Std Error: 0.00		
	Seq primer: -40ml3 fwd. ET from Amersham		
	High quality sequence stop: 359.		
FEATURES			
source	1. .498		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1091276"		
	/clone_id="jia bone marrow stroma"		

BASE COUNT	117 a	137 c	82 g	162 t
ORIGIN	<pre> /sex="mixed" /tissue_type="bone marrow stroma" /dev_stage="mixed" /lab_host="X11-Blue MR"/SOLR" /ncbi_vector="blueprint; Site:1: EcorI; Site:2: XhoI; RNA made from human bone marrow stroma, cDNA made by oligo-dt priming. Directionally cloned. Size selected for average insert size >0.5 kb. Library supplied by Dr. Libin Jia (NHRRI)." </pre>			
Query Match	35.2%	Score 484.8;	DB 36;	Length 498;
Best Local Similarity	98.2%;	Pred. No. 2.5e-109;		
Matches 489;	Conservative	0;	Mismatches 9;	Indels 0;
			Gaps 0;	
QY	799	GGGATGGAAGAACGACGCTGGACATAGAAATAGACCCCTGGGGCTGAGAGACGGCATGGAG	858	
Db	498	GGGATGGAAGAACGACGCTGGACATAGAAATAGACCCCTGGGGCTGAGAGACGGCATGGAG	439	
QY	859	TACCCCTTATTTGAGAGAGGTAGCCCTCACGATGAGGACCTTGAGATTTCAGGTTCC	918	
Db	438	TACCCCTTATTTGAGAGAGGTAGCCCTCACGATGAGGACCTTGAGATTTCAGGTTCC	379	
QY	919	CGAATCAAAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGAGATGATTGTACACAAAT	978	
Db	378	CGAATCAAAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGAGATGATTGTACACAAAT	319	
QY	979	GTGACAGTCTCATTTAGTTGATGACACTGGTGGCTTTGAGATGATATTTACTACTTGGAT	1038	
Db	318	GTGACAGTCTCATTTAGTTGATGACACTGGTGGCTTTGAGATGATATTTACTACTTGGAT	259	
QY	1039	GGTCAAGAGGTACATATTTCCGGGATTAAGATCACAGGCCAGAGACCAANTANTTGAAG	1098	
Db	258	GGTCAAGAGGTACATATTTCCGGGATTAAGATCACAGGCCAGAGACCAANTANTTGAAG	199	
QY	1099	AAAGGGGAAGGGCTCCCACTTTGACACAACAACATATCAAGGGCTCTTTGATTAATCACT	1158	
Db	198	AAAGGGGAAGGGCTCCCACTTTGACACAACAACATATCAAGGGCTCTTTGATTAATCACT	139	
QY	1159	TTTGATGTGAGATTTTCCAAAAGAACAGTTAACAGAGAGAACGAGACGTATCAACAG	1218	
Db	138	TTTGATGTGAGATTTTCCAAAAGAACAGTTAACAGAGAGAACGAGACGTATCAACAG	79	
QY	1219	CTACGAAACAAGGGGTCACTGTCGAGAGGTATATCAATGCACTCAAGAGATTTGAGAGTGA	1278	
Db	78	CTACGAAACAAGGGGTCACTGTCGAGAGGTATATCAATGCACTCAAGAGATTTGAGAGTGA	19	
QY	1279	ATMAAATTGGACTTTGTT 1296		
Db	18	ATMAAATTGGACTTTGTT 1		
RESULT 6				
LOCUS	AA581995	EST		05-SEP-1997
DEFINITION	nm56d05.s1 NCI_CGAP GC5 Homo sapiens cDNA clone IMAGE:1085961 3'			
ACCESSION	AA581995			
VERSION	AA581995.1			
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
AUTHORS	Eukheria; Primates; Catarrhini; Hominiidae; Homo.			
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.			
JOURNAL	National Cancer Institute, Cancer Genome Anatomy Project (CGAP),			
COMMENT	Tumor Gene Index			
	Unpublished (1997)			
	Contact: Robert Strausberg, Ph.D.			
	Tel.: (301) 496-1550			

Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Stratagene, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 .DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 www-bio.llnl.gov/bdrrp/image/image.html

Seq primer: -40m13 fwd. ER from Amersham
 High quality sequence stop: 344.

FEATURES

Location/Qualifiers

1. 488

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1085961"

/clone_1lb="NCI-CGAP-GC5"

/tissue_type="germ cell tumor"

/lab_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Mixed

germ cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGCAG

3' 3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'

Average insert size: 0.7 kb.

135 a 135 c 80 g 158 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 35.0%; Score 481.2; DB 35; Length 488;

Matches 483; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

811 CGAAGCTGAGTAGAATAAGACCTGGGTGAGACAGCGCATGAGTACCCCTTATT 870

488 CGAAGCTGAGTAGAATAAGACCTGGGTGAGACAGCGCATGAGTACCCCTTATT 429

871 GGAGAGGTAGCTCAGCTGATGGGAGCCTGGAGATTACGGTCCGAATCAAGTT 930

428 GGAGAGGTAGCTCAGCTGATGGGAGCCTGGAGATTACGGTCCGAATCAAGTT 369

931 GTCAGAGCACCATTATTGTAAGAGAGAGATGATTGTCACAAATGCAAGTCTCA 990

368 GTCAGAGCACCATTATTGTAAGAGAGAGATGATTGTCACAAATGTCACATCTCA 309

991 TTAGTGAAGTCACTGGTGGCTTTGAGATGATATTACTGCTGATGCTCAAGGTA 1050

308 TTAGTGAAGTCACTGGTGGCTTTGAGATGATATTACTGCTGATGCTCAAGGTA 249

1051 CATATTCCCGGATTAAGATCAGACAGCCAGGAGCCGCAATANTGAAAGAGGAAAGG 1110

248 CATATTCCCGGATTAAGATCAGACAGCCAGGAGCCGCAATANTGAAAGAGGAAAGG 189

1111 CTCGCCAAGTCTTGACAACAATATCAAGGCTCTTTGATATCACTTTTGTGATGAT 1170

188 CTCGCCAAGTCTTGACAACAATATCAAGGCTCTTTGATATCACTTTTGTGATGAT 129

1171 TTTCGAAAAGACGTTAAACAGAGAGAGCGAGAGAGATATCAACAGCTACTGAAACAA 1230

128 TTTCGAAAAGACGTTAAACAGAGAGAGCGAGAGAGATATCAACAGCTACTGAAACAA 69

1231 GGCTGCTCAGAGAGATATCAATGATGCTGCAAGGATATTTGAGAGTAAATTTGAC 1290

68 GGCTGCTCAGAGAGATATCAATGATGCTGCAAGGATATTTGAGAGTAAATTTGAC 9

1291 TTTGTTTA 1298

8 TTTGTTTA 1

RESULT 7

N93316/c

LOCUS

N93316

482 bp

mRNA

EST

20-AUG-1996

DEFINITION

z69603.s1 Soares_fetal_lung_NbHL19W Homo sapiens CDNA clone
 IMAGE:308860 3' similar to WP:TI5H9.1 CE01664 DNAJ ; mRNA
 sequence.

ACCESSION

N93316

VERSION

GI:1265625

KEYWORDS

EST.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 482)

REFERENCE

AUTHORS

Chilcote, S., Lennon, G., Becker, M., Bonaldo, M.F., Chiapelli, B.,

Hillier, L., Lennon, G., Dubuque, T., Favell, A., Gish, W.,

Hawkins, M., Hultman, M., Kucaba, T., Lacy, M., Le, M., Le, N.,

Mardis, E., Moore, B., Morris, M., Parsons, J., Prange, C., Rifkin, L.,

Rohlfing, T., Schellenberg, K., Soares, M.B., Tan, F., Thierry-Mieg, J.,

Trevaskis, E., Underwood, K., Wohlmann, P., Waterston, R., Wilson, R.,

and Marra, M.

Generation and analysis of 280,000 human expressed sequence tags

Genome Res. 6 (9), 807-828 (1996)

97044478

CONTACT: Willson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Possible reversed clone: similarity on wrong strand

Insert Length: 1849 Std Error: 0.00

Seq primer: mob.REGA+ET

High quality sequence stop: 470.

Location/Qualifiers

1. 482

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:308860"

/clone_1lb="Soares_fetal_lung_NbHL19W"

/dev_stage="19 weeks"

/lab_host="DH10B (ampicillin resistant)"

/note="Organ: lung; Vector: pT73D (Pharmacia) with a

modified polylinker; Site: 1: Not I; Site: 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer.

15'-TGTTACCAATCTGAGTGGAGGCGCGCAATTTTCTTTTCTTTTCTTTT-3',

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot - 5. Library constructed by Bento

Soares and M. Fatima Bonaldo. This library was constructed

from the same fetus as the fetal heart library, Soares

fetal heart NbHL19W."

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 34.7%; Score 476.8; DB 25; Length 482;

Matches 478; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

817 CTGGAAGTAGAATAGAGCTGGGTGAGAGAGCGCATGAGTACCCCTTATTGGAGAA 876

482 CTGGAAGTAGAATAGAGCTGGGTGAGAGAGCGCATGAGTACCCCTTATTGGAGAA 423

877 GGTGAGCTCAGCTGATGGGAGGCTGAGATTACGGTCCGAATCAAGTTGTCAAG 936

422 GGTGAGCTCAGCTGATGGGAGGCTGAGATTACGGTCCGAATCAAGTTGTCAAG 363

937 CACCCAAATTTGAAGAAGAGAGAGATGATTGTACAAATGTCAGCTCATTTAGTT 996

362 CACCCAAATTTGAAGAAGAGAGAGATGATTGTACAAATGTCAGCTCATTTAGTT 303

QY 997 GAGTCACTGGTTGGCTTGGATGATATTAATCTGCTTGGATGGTGCACAGGTACATATT 1056
|||||
Db 302 GAGTCACTGGTTGGCTTGGATGATATTAATCTGCTTGGATGGTGCACAGGTACATATT 243
QY 1057 TCCCGGGAATAGATCACCAGGCCAGACGCAANTANTGAAGAAAGGGAGGCTCCCC 1116
|||||
Db 242 TCCCGGGAATAGATCACCAGGCCAGACGCAANTANTGAAGAAAGGGAGGCTCCCC 183
QY 1117 AACTTTGCAACAACATATTCAGGGCTCTTTGATTAATCACTTTTGGATGGATTTTCCA 1176
|||||
Db 182 AACTTTGCAACAACATATTCAGGGCTCTTTGATTAATCACTTTTGGATGGATTTTCCA 123
QY 1177 AAGAAGCTTAACAGAGAGAGCGGAGAGATATCAACAGCTCTGTAACAAAGGCTCA 1236
|||||
Db 122 AAGAAGCTTAACAGAGAGAGCGGAGAGATATCAACAGCTCTGTAACAAAGGCTCA 63
QY 1237 GTGCAAGATATACAAATGACATCTGCAAGATATTTGAGATGATTAATTAATTTGACTTTGT 1296
|||||
Db 62 GTGCAAGATATACAAATGACATCTGCAAGATATTTGAGATGATTAATTAATTTGACTTTGT 3
QY 1297 TA 1298
||
Db 2 TA 1
RESULT 8
A1819362/c 791 bp mRNA EST 24-AUG-1999
LOCUS w61b11.x1 Soares NSF_P8_9W_OT_PA_P.S1 Homo sapiens cDNA clone
DEFINITION IMAGE:2369565.3 similar to WP:T15H9.1 CE01664 HYPOTHETICAL. 39.9 KD
PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ;, mRNA sequence.
ACCESSION A1819362 GI:5438441
VERSION A1819362.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 791)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jun 5, 1998 this sequence version replaced gi:3189482.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Seq primer: -400P from G1Dco
High quality sequence stop: 469.
Location/Qualifiers
1. 791
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2369565"
/clone_id="Soares_NSF_P8_9W_OT_PA_P.S1"
/lab_host="DH10B"
/note="Organ: pooled; Vector: pT73D-Pac (Pharmacia) with
a modified polylinker; Site_1: Not I; Site_2: Eco RI;
Equal amounts of plasmid DNA from five normalized
libraries were mixed, and ss circles were used as tracer in
Following HAP purification, this DNA was used as tracer in
a subtractive hybridization reaction. The driver was
PCR-amplified cDNAs from pools of 5,000 clones made from
the same 5 libraries. The pools consisted of the following
libraries and clones: Soares NBHSF pool 1:
309384-310919, 323208-323895 Soares NB2HP pool 1:
145032-147335, 147720-148103, 148872-149255, 15002 -
150407, 151176-152327 Soares NB2HP-9W pool 1:
758280-760583, 772104-774407 Soares NBHPA pool 1:
304776-306311, 320136-322823, 326280-326663 Soares NBHOT

BASE COUNT 234 a 186 c 114 g 254 t 3 others
ORIGIN
Query Match 33.1%; Score 455.6; DB 61; Length 791;
Best Local Similarity 95.0%; Pred. No. 4.5e-102;
Matches 478; Conservative 0; Mismatches 24; Indels 1; Gaps 1;
QY 874 GAAGTGAGCCCTCAGCTGATGGAGCCCTGGAGATTTTACGGTTCGAATCAAGTGTGC 933
|||||
Db 791 GAAGTGAGCCCTCAGCTGATGGAGCCCTGGAGATTTTACGGTTCGAATCAAGTGTGC 732
QY 934 AAGCACCACATATTTTGAAGAGAGAGATGATTTTGTACACAATGTGACATCTCATTA 993
|||||
Db 731 AAGCACCACATTA-TTGAAAGAGAGAGATGATTTTGTACACAATGTGACATCTCATTA 673
QY 994 GTTGAGTCACTGGTGGCTTTGAGATGATATTAATCACTTGGATGGTGCACAAAGTACT 1053
|||||
Db 672 GTTGAGTCACTGGTGGCTTTGAGATGATATTAATCACTTGGATGGTGCACAAAGTACT 613
QY 1054 ATTTCCCGGATATGATATCAGCCAGGCGCAANTANTGAAGAAAGGGAGGCTC 1113
|||||
Db 612 ATTTCCCGGATATGATATCAGCCAGGCGCAANTANTGAAGAAAGGGAGGCTC 553
QY 1114 CCCAATTTGACACACATATTCAGGGCTCTTTGATTAATCACTTTGATGTGATTTT 1173
|||||
Db 552 CCCAATTTGACACACATATTCAGGGCTCTTTGATTAATCACTTTGATGTGATTTT 493
QY 1174 CCNAAAGACATTTTACAGAGAGAGACGAGAGAGATTAACACACTACTGAAACAAGG 1233
|||||
Db 492 CCNAAAGACATTTTACAGAGAGAGACGAGAGATTAACACACTACTGAAACAAGG 433
QY 1234 TCAGTGCAAGATATACATGACTGCAAGATATTTGAGATGATTAATTAATTTGACTTT 1293
|||||
Db 432 TCAGTGCAAGATATACATGACTGCAAGATATTTGAGATGATTAATTAATTTGACTTT 373
QY 1294 GTTTAAATTAAGTGAATAAGCATATTTATATCTGCAGAGTTTGTGTGTGTTT 1353
|||||
Db 372 GTTTAAATTAAGTGAATAAGCATATTTATATCTGCAGAGTTTGTGTGTGTTT 313
QY 1354 GTTTTATTTTCAATATGCAACT 1376
|||||
Db 312 GTTTTATTTTCAATATGCAACT 290
RESULT 9
AAB37534/c 767 bp mRNA EST 18-MAR-1998
LOCUS oes1e05.s1 NCI-CCAP_Pr25 Homo sapiens cDNA clone IMAGE:1410176
DEFINITION similar to WP:T15H9.1 CE01664 DNMF; contains Alu repetitive
element; mRNA sequence.
ACCESSION AAB37534 GI:2912733
VERSION AAB37534.1
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 767)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
TITLE Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jan 17, 1998 this sequence version replaced gi:1900092.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Suzanne L. Topalian, M.D., Robert K. Bright,
Ph.D.
cDNA Library Preparation: Stratagene, Inc.
cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

Insert Length: 1639 Std Error: 0.00
 Seq primer: -40ml3 fwd. Ex from Amersham
 High quality sequence stop: 487.
 Location/Qualifiers

FEATURES
 source
 1767
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:1410176"
 /clone_1id="NCI CGAP_P25"
 /tissue_type="epithelium (cell line)"
 /lab_host="SOAR (kanamycin resistant)"
 /note="Organ: prostate; Vector: Bluescript SK-; Site_1:
 EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:
 Oligo dt. Normal prostate epithelial cell line (HPV
 immortalized). 5' adaptor sequence: 5' GAATTCGGCAGAG 3'
 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
 Average insert size: 1.1 kb."
 BASE COUNT 227 a 177 c 114 g 249 t
 ORIGIN

Query Match 32.4%; Score 445.4; DB 39; Length 767;
 Best Local Similarity 96.1%; Pred. No. 1.5e-99;

Matches 466; Conservative 0; Mismatches 18; Indels 1; Gaps 1;

892 GATGGGAGCTGAGATTTACGTTCCGAAATCAAGTTGTCAAGCACCCAAATATTGAA 951
 767 GATGGGAGCTGAGATTTACGTTCCGAAATCAAGTTGTCAAGCACCCAAATATTGAA 708
 952 AGGAGAGAGATGATTTGTACAAATGTGACATCTAGTTGAGTACTGCTGGC 1011
 707 AGGAGAGAGATGATTTGTACAAATGTGACATCTAGTTGAGTACTGCTGGC 648
 1012 TTGAGATGATTTACTACTGTTGATGTCACAAAGGTACATATTTCCGGATAGATC 1071
 647 -TTGAGATGATTTACTACTGTTGATGTCACAAAGGTACATATTTCCGGATAGATC 589
 1072 ACCGAGCAGAGAGGAGTANTGGAAGAGGAGGCTCCCAACTTTGACACAC 1131
 588 ACCGAGCAGAGAGGAGTANTGGAAGAGGAGGCTCCCAACTTTGACACAC 529
 1132 AATATCAAGGCTCTTGTGATATCACTTTGATGTCGATTTCCAAAGAAGATTACA 1191
 528 AATATCAAGGCTCTTGTGATATCACTTTGATGTCGATTTCCAAAGAAGATTACA 469
 1192 GAGGAGCGAGAGAGTATCAACACACTGTAACAAAGGTCAGTCAGAAAGATTAC 1251
 468 GAGGAGCGAGAGAGTATCAACACACTGTAACAAAGGTCAGTCAGAAAGATTAC 409
 1252 AATGAGCTGCAAGGATTTAGAGTGAATTAATTTGACTTTTAAATAAGTAA 1311
 408 AATGAGCTGCAAGGATTTAGAGTGAATTAATTTGACTTTTAAATAAGTAA 349
 1312 AGGATATTTATTTATCTGCAAGGTTTTTTTGTGTTTGTGTTTATTTCAATATG 1371
 348 AGGATATTTATTTATCTGCAAGGTTTTTTTGTGTTTGTGTTTATTTCAATATG 289
 1372 CAAGT 1376
 288 CAAGT 284

RESULT 10
 AM323622 556 bp mRNA EST 26-JAN-2000
 LOCUS uc74e06.y1 NCI-CGAP_Mam1 Mus musculus cDNA clone IMAGE:2648290 5'
 DEFINITION similar to WP:115H9.1 CE01664 HYPOTHETICAL 39.9 KD PROTEIN T15H9.1
 IN CHROMOSOME II PRECURSOR ; mRNA sequence.

ACCESSION AM323622
 VERSION AM323622.1 GI:6757647
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus

REFERENCE Eukaryotes: Metazoa: Chordata: Craniata: Vertebrata: Mammalia;
 Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: Mus.
 1 (bases 1 to 556)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)

JOURNAL On Jun 22, 1998 this sequence version replaced gi:3246929.
 COMMENT Other-ESTs: uc74e06.x1
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbp/image/image.html

NCI:1028742
 Seq primer: -40RP from Gibco
 High quality sequence stop: 304.
 Location/Qualifiers

FEATURES

source

1.556
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:2648290"
 /clone_1id="NCI CGAP_Mam1"
 /tissue_type="tumor, biopsy sample"
 /dev_stage="3 months, virgin"
 /lab_host="DH10B"
 /note="Organ: mammary; Vector: pCMV-Sport6; Site_1: SalI;
 Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
 Library constructed by Life Technologies. Investigator
 providing samples: Gilbert Smith, NIH"
 BASE COUNT 156 a 110 c 161 g 129 t
 ORIGIN

Query Match 32.4%; Score 445.2; DB 80; Length 556;
 Best Local Similarity 87.7%; Pred. No. 1.5e-99;

Matches 486; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

513 TTCACACTCTTTGGGATTTGTTGATGTTGGAGGAGCCCTCGCAGCAGACAG 572
 1 TTCACACTCTTTGGGATTTGTTGATGTTGGAGGAGCCCTCGCAGCAGACAG 60
 573 AATATTCAGAGAGGATATTTATTTAGATCTAGAAGTCACTTTGGAAGATATA 632
 61 GAATATTCAGAGAGGATATTTATTTAGATCTAGAAGTCACTTTGGAAGATATA 120
 633 TCGAGAAATTTTGGAGAGTATGAAGAAACACTGTGGCAAGGAGCTCTGGCA 692
 121 CGAGGAAATTTTGGAGAGTATGAAGAAACACTGTGGCAAGGAGCTCTGGCA 180
 693 ACGAAGTCAATTTGCGCAGAGATGCGACACCCAGTGGGCGCTGGCTTCCA 752
 181 ACGAAGTCAATTTGCGCAGAGATGCGACACCCAGTGGGCGCTGGCTTCCA 240
 753 AATGACCCAGAGGTGTCTGCGAGAAATGCCCTAATGTCAAACTAGTGAAGAAG 812
 241 AATGACCCAGAGGTGTGTGAGAGTGTGGAGAGTCCCTAATGTCAAACTAGTGAAGAAG 300
 813 AACGGTGAAGTGAAGAAATGAGCCCTGGGTGAGAGACCGCATGAGATCCCTTTATGG 872
 301 AACCTAGAGTGAAGTGAAGCCCTGGGTGAGATGATGAGATGAGATCCCTTTATGG 360

QY	873	AAAGGTACGCTCAGCGGATGGGAGCGCTGAGATTTAGCTTCGCAATCAAGGTGT	932
DB	361	AGAAAGTGAACCGGACATTCGATGCTGAACCTCGAGACTTACGGTTCGATCAAAAGCTGT	420
QY	933	CAAGCACCATAATTTTGAAGAGAGAGAGATGATTTGTACACAATATGTACATCTCAAT	992
DB	421	CAAGCTCCGGATATTTTGATGATGAGAGGGGATGACCTGTACACAATATGTACCGTCTCACT	480
QY	993	AGTTATGTCACCTGGTGGCTTTGACATGATATTTACTCATCTTGATGGTCACAGGTACA	1052
DB	481	GGTTCAGGCTCTCGGGTGGCTTTGATATGTGACATACTGAATCGATGTGTACACAGCTTGG	540
QY	1053	TATTTCCCGGATA	1066
DB	541	TATTCGCCGTGATA	554
RESULT	11		
LOCUS	AA596749		
DEFINITION	AA596749	531 bp	mRNA
ACCESSION	AA596749		
VERSION	AA596749.1	GI:2412184	
KEYWORDS	EST.		
SOURCE	house mouse.		
ORGANISM	Mus musculus		
REFERENCE	Euthyria; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
AUTHORS	1 (bases 1 to 531) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.		
TITLE	The Washu-HMI Mouse EST Project		
JOURNAL	Unpublished (1996)		
COMMENT	On Sep 12, 1996 this sequence version replaced gi:1315918. Contact: Marra M/Mouse EST Project Washu-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email: mouseest@watson.wustl.edu This clone is available royalty-free through LINT; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:566784 Seq primer: -28m13 rev1 ET from Amersham High quality sequence stop: 445. Location/Qualifiers 1..531 /organism="Mus musculus" /db_xref="taxon:10090" /clone="IMAGE:1002568" /clone_lib="Stratagene mouse Tcell 937311" /tissue_type="Tcell" /dev_stage="M30 CD4+ cells" /lab_host="SODR (kanamycin resistant)" /note="Organ: blood; Vector: plusscript SK-; Site: 1; EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer: 0150 dt. M30 CD4+ cells. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCGAGAG 3' -3' adaptor sequence: 5' CTCGAGCTTTTCTTTTCTTTT 3'."		
BASE COUNT	162 a	108 c	154 g
ORIGIN	107 t		
Query Match	32.0%	Score 441;	DB 35; Length 531;
Best Local Similarity	89.3%;	Pred. No. 1.6e-98;	
Matches 474;	Conservative	0;	Mismatches 57;
		Indels	0;
		Gaps	0;

Oy	713	AAGGATATGGGACACCCACAGCTGGGCGCCGCGGGCGCTTCCAAATGATGACCCAGGAGGTGCT	772
Db	1	AAGGATGAAACCCACACAGCTGGGACCGGGACGGCTTCCAAATGATGACCCAGGAGGTGCT	60
Oy	773	GCACGCAATCCCTTAATGTCAAACTAGTAATGAAGAACAAGCAACCGCTGGAGATAGAAATAG	832
Db	61	GTGACAGCTGCCCTTAATGTCAAACTAGTAATGAAGAACAAGCAACCTAGTAATGAATAG	120
Oy	833	AGCGTGGGTGAGAGACGGCGATGAGTACCCCTTTATTTGGAAGAGTGAAGCTTCAGCTG	892
Db	121	AGCGTGGGTGAGAGATGAGTGAATGAGTACCCCTTTATTTGGAAGAGTGAAGCTTCAGCTG	180
Oy	893	ATGGGGAGCCTGGAGATTACAGCTTCCGAATCAAAAGCTGTCAACACCCATATTATTTAA	952
Db	181	ATGGGGAGCCTGGAGACTTACAGCTTCCGAATCAAAAGCTGTCAACACCCGATATTATTTAGA	240
Oy	953	GGAGAGAGATGATTTGTATACACAAATGATACAGTCTCATTTAGTTGAGTCACTAGTGGCT	1012
Db	241	GGAGAGGGGATGACCTGTATACAAAAATGTGACCGTCTCACTAGTGGTGTGGTGGCT	300
Oy	1013	TTGAGATGATATTACTCACTTGGATGTCACACAGGTACATATTTTCCGGGATTAAGTCA	1072
Db	301	TTGAGATGAGCAATACATCACTGGAGTGCACAAAGGCTCATATTTCCCGGGACAAAGTCA	360
Oy	1073	CCAGGCCAGAGGAGGANTTANTGGGAAGAAAGGGGACGGCTCCCAACTTTGACACACACA	1132
Db	361	CCAGGCCAGAGGCCAAGCTGTGGAAAGAAAGGGGAGGGCTCCCAACTTTGATATACATA	420
Oy	1133	ATATCAAGGCGCTTTTGATATATCATCTTTTGATGTGATTTTCCAAAGAACATTAACAG	1192
Db	421	ACATCAAGGCGCTTTTGATATATCATCTTTTGATGTGATTTTCCAAAGAACATGTGACAG	480
Oy	1193	AGGAAGCGAAGAAGGTATCAAAACAGCTACTGTAACAAAGGCTCAAGTGAAGA	1243
Db	481	AGGAAGCGAAGAAGGTATCAAAACAGCTTTCAACAAAGGCGCGGTGAGA	531
RESULT	12		
LOCUS	AI057091/c		
DEFINITION	AI057091	750 bp	mRNA
ACCESSION	AI057091		EST
VERSION	AI057091		01-OCT-1998
KEYWORDS	AI057091.1	GI:3330967	
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;		
AUTHORS	Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 750)		
JOURNAL	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .		
COMMENT	National Cancer Institute, Cancer Genome Anatomy Project (CGAP)		
	Unpublished (1997)		
	On May 18, 1995 this sequence version replaced gi:811070.		
	Contact: Robert Strausberg, Ph.D.		
	Tel: (301) 496-1550		
	Email: Robert.Strausberg@nih.gov		
	This clone is available royalty-free through LML; contact the		
	IMAGE Consortium (info@image.llnl.gov) for further information.		
	Insert Length: 862 Std Error: 0.00		
	Seq primer: -40m13 fwd. ET from Amersham		
	High quality sequence stop: 445.		
FEATURES	Location/Qualifiers		
source	1..750		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:1676202"		
	/clone_lib="Soares_total_fetus_NB2HR8_9w"		
	/dev_stage="8-9 weeks"		
	/lab_host="DH10B"		

RESULT 14
 A1800795/c
 LOCUS
 DEFINITION
 A1800795 726 bp mRNA EST 19-DEC-1999
 w913c05.x1 Soares.NSF_F8_9W_OT_PA_P_S1 Homo sapiens cDNA clone
 IMAGE:2364968 3' similar to WP:115H9.1 CE01664 HYPOTHETICAL 39.9 KD
 PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR ; mRNA sequence.
 A1800795
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 726)
 NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished (1997)
 On Feb 22, 1999 this sequence version replaced gi:4283012.
 COMMENT
 Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1350
 Email: Robert.Strausberg@nih.gov
 This clone is available royalty-free through LNL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.
 Insert length: 1027 Std Error: 0.00
 Seq primer: -400P from Gdbco
 High quality sequence stop: 445.
 Location/Qualifiers
 1. 726
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2364968"
 /clone_lib="Soares.NSF_F8_9W_OT_PA_P_S1"
 /lab_host="DH10B"
 /note="Organ: pooled: Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker: Site.1: Not I; Site.2: Eco RI;
 Equal amounts of plasmid DNA from five normalized
 libraries were mixed, and ss circles were made in vitro.
 Following HAP purification, this DNA was used as tracer in
 a subtractive hybridization reaction. The driver was
 PCR-amplified cDNAs from pools of 5,000 clones made from
 the same 5 libraries. The pools consisted of the following
 libraries and clones: Soares NBHSR pool 1:
 309384-310919, 323208-325895 Soares NB2HP pool 1:
 145032-147335, 147720-148103, 148872-149255, 15002 -
 150407, 151176-152327 Soares NB2HF8-9W pool 1:
 758280-760583, 772104-774407 Soares NBHFA pool 1:
 304776-306311, 320136-322823, 326280-326663 Soares NBHOR
 pool 1: 723720-726407, 739080-740999 Subtraction by Bento
 Soares and M. Fatima Bernaldo."
 BASE COUNT 215 a 164 c 104 g 241 t 2 others
 ORIGIN

Query Match 30.3%; Score 417; DB 60; Length 726;
 Best Local Similarity 96.8%; Pred. No. 1.5e-92;
 Matches 423; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

QY 940 CCAATATTGAAAGGAGAGAGATTTGTACCAAAATGAGAGCTCATTTAGTTGAG 999
 |||
 Db 726 CCCAAATATTGAAAGGAGAGAGATTTGTACCAAAATGAGAGCTCATTTAGTTGAG 667
 QY 1000 TCACTGGTGGCTTTGAGATGATATTACTACCTGGATGGTCCAAAGTACATATTCC 1059
 |||
 Db 666 TCACTGGTGGCTTTGAGATGATATTACTACCTGGATGGTCCAAAGTACATATTCC 607
 QY 1060 CGGAGTAAGTACACAGGCCAGAGCGAANTANTGAGAGAAAGGGGAGGCTCCCAAC 1119
 |||
 Db 606 GGGGATTAAGTACACAGGCCAGAGCGAGCTATGAGAAAGGAGGAGGCTCCCAAC 547
 QY 1120 TTGACACAAACATATCAAGGGCTCTTTGATATCACTTTTGTATGATTTGCCAAA 1179
 |||

Db 546 TTGACACAAACAAATATCAAGGGCTCTTTGATTAATCACTTTTGTATGATTTGCCAAA 487
 QY 1180 GAACAGTTAAACAGAGAGAGAGAGAGGATTTCAACAGCTACTGAAACAGGGTCACTG 1239
 |||
 Db 486 GAACAGTTAAACAGAGAGAGAGAGAGGATTTCAACAGCTACTGAAACAGGGTCACTG 427
 QY 1240 CAGAAGTATCAATGAGCTCAAGGATATGAGAGTAAATTAATGACTTTGTTTAA 1299
 |||
 Db 426 CAGAAGTATCAATGAGCTCAAGGATATGAGAGTAAATTAATGACTTTGTTTAA 367
 QY 1300 AATAAGTAAATAGCGAATATTTATCTGCAAGGTTTTTTGTGTGTTTTGTTTTT 1359
 |||
 Db 366 AATAAGTAAATAGCGAATATTTATCTGCAAGGTTTTTTGTGTGTTTTGTTTTT 307
 QY 1360 ATTTCAATATGCAAGT 1376
 |||
 Db 306 ATTTCAATATGCAAGT 290

RESULT 15
 AM122551/c
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Mus musculus
 house mouse.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 480)
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 97044477
 COMMENT
 On Feb 22, 1999 this sequence version replaced gi:4283549.
 Contact: Chin, H
 National Institute of Mental Health
 6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD
 20892-9643, USA
 Tel: 301 443 1706
 Fax: 301 443 9890
 Email: mbest@mail.nih.gov
 Oligo-dT track not found, Not I site shown in beginning of sequence
 is likely internal to the message. cDNA library Preparation: M.B.
 Soares lab clone distribution: NIH BMAP cDNA clones will be made
 available by the means that is soon to be determined. When NIH
 determines the means for distribution of the BMAP cDNA clones, this
 record will be updated accordingly when that means is determined.
 Seq primer: M13 Forward
 POLYA-No.

FEATURES
 source
 Location/Qualifiers
 1. 480
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="UI-M-BH2.2-sox-d-09-0-UI"
 /clone_lib="NIH_BMAP_M.S3.2"
 /dev_stage="27-32 days"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker: Site.1: Not I; Site.2: Eco RI; The
 NIH BMAP M.S3.2 library is a subtracted library of a
 series, ultimately derived from a mixture of individually
 tagged normalized libraries from ten regions of the mouse
 brain (cerebellum, brain stems, olfactory bulbs,
 hypothalamus, cortex, amygdala, basal ganglia, pineal
 gland, striatum, hippocampus) after a series of
 subtractions to reduce the representation of cDNAs from
 which ESTs had already been generated. The following
 serially subtracted libraries were generated in this

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:57:25 ; Search time 40.12 Seconds
(without alignments)
271.757 Million cell updates/sec

Title: US-09-501-714-1
Perfect score: 1883
Sequence: 1 MAPQMLSTFCLLLYLIGAV.....IKQLLKQSGVQKVTNGLOGY 358

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues
Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SwissProt_38:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1037.5	55.1	355 1	YRY1_CAEEL
2	549.5	29.2	418 1	DNU2_ALIPO
3	548.5	29.1	419 1	DNUH_ARATH
4	538.5	28.6	413 1	DNUH_CUCSA
5	526.5	28.0	417 1	DNUH_ARATH
6	519.5	27.6	397 1	DNU2_ALIPO
7	515.5	27.4	397 1	DNU2_HUMAN
8	515.5	27.4	397 1	DNU2_MOUSE
9	510	27.1	409 1	MAS5_YEAST
10	469	24.9	375 1	DNAJ_BRNOV
11	450.5	23.9	375 1	DNAJ_ECOLI
12	447.5	23.8	378 1	DNAJ_SALTY
13	446.5	23.7	375 1	DNAJ_ACTAC
14	441.5	23.4	369 1	DNAJ_LEPIN
15	441	23.4	376 1	DNAJ_METH
16	440	23.4	367 1	DNAJ_COXBU
17	437	23.2	340 1	HS41_HUMAN
18	435.5	23.1	369 1	DNAJ_NTEU
19	428	22.7	334 1	DNUJ_ROME
20	425.5	22.6	384 1	DNAJ_RHOCA
21	419.5	22.3	377 1	DNAJ_HABDU
22	419	22.3	374 1	DNAJ_CIOAB
23	417	22.1	369 1	DNAJ_STRCO
24	415	22.0	371 1	DNAJ_HAEIN
25	409.5	21.7	389 1	DNAJ_FRATU
26	405.5	21.5	379 1	DNAJ_METMA
27	403	21.4	379 1	DNAJ_LEGPN
28	400	21.2	377 1	DNAJ_BRAJA
29	398	21.1	370 1	DNAJ_RICPR
30	390.5	20.7	379 1	PSI_SCHPO
31	385.5	20.5	379 1	DNAJ_STPAU
32	385	20.4	364 1	DNAJ_BORBU
33	374.5	19.9	395 1	DNAJ_MYCTU
34	369	19.6	459 1	KDJ1_YEAST

35	366	19.4	379 1	DNAJ_IACIA	P35514 lactococcus
36	364.5	19.4	280 1	DNAJ_THETH	O56237 thermus agu
37	362.5	19.3	352 1	SIS1_YEAST	P25234 saccharomyc
38	355	18.9	372 1	DNAJ_BACSU	P17631 bacillus su
39	354	18.8	528 1	YNH7_YEAST	P53940 saccharomyc
40	351	18.6	404 1	SCU1_YEAST	P25303 saccharomyc
41	348.5	18.5	370 1	DNAJ_ERYRH	O05646 erysipelo
42	345	18.3	352 1	DNAJ_STRPN	P95830 streptococ
43	344.5	18.3	389 1	DNAJ_MYCPE	P47265 mycoplasma
44	344.5	18.3	389 1	DNAJ_MYCLE	O02605 mycobacteri
45	337.5	17.9	234 1	DNAJ_RHILE	O33529 rhizobium 1

ALIGNMENTS

RESULT	ID	YRY1_CAEEL	STANDARD	PRT	355 AA.
AC	O10005				
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	HYPOTHETICAL 39.9 KD PROTEIN T15H9.1 IN CHROMOSOME II PRECURSOR.				
GN	T15H9.1.				
OS	Caenorhabditis elegans.				
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;				
OC	Rhabditina; Rhabditioidea; Rhabditidae; Peloderinae; Caenorhabditis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-BRISTOL NZ;				
RA	Gardner A.;				
RL	Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.				
CC	-1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
DR	EMBL; Z47356; CA87414.1; -.				
DR	HSSP; P08622; IXL.				
DR	WORMPEP; T15H9.1; CE01664.				
DR	PROSITE; PS00636; DNAJ_1; 1.				
DR	PROSITE; PS00706; DNAJ_2; 1.				
DR	PRAM; PF00226; DnaJ_1.				
DR	PRAM; PF01356; DnaJ_C; 1.				
KW	Hypothetical protein; Chaperone; Signal.				
FT	SIGNAL 1 21				
FT	CHAIN 22 355				
FT	DOMAIN 22 91				
FT	DOMAIN 97 103				
FT	DOMAIN 116 123				
FT	DOMAIN 355 AA; 39861 MW; 95FAD8551D9CC2 CRC64;				
SQ	SEQUENCE				
Query Match	55.1%;	Score 1037.5;	DB 1;	Length 355;	
Best Local Similarity	58.0%;	Pred. No. 3.9e-78;			
Matches 203;	Conservative 54;	Mismatches 84;	Indels 9;	Gaps 5;	
OY	13 LLYLVIGAVIA-----GRPFYKILGYPRASISIDIKKAYRKLAQLHPDPNPDDPOAQEKFQ 68				
DB	8 LTVLASSIVAAVEEGRDYKILGVAKANNANQIKAYRKLAKEIHPDNQDDMANEKFKQ 67				
OY	69 DLGAAYEVLDSSEKRRKQDYTGEEGL-KDGHQSSHG-DIESHFPGDGFEMFGGTPROODR 126				
DB	68 DLSSAYEVLDSSEKRRKQDYTGEEGL-KDGHQSSHG-DIESHFPGDGFEMFGGTPROODR 125				
OY	127 NIPGSDIIVLEVTLEEVYVGNFVVRNKPFAVROAPGRKCKCROEMRTQIGPGRFQ 186				

Best Local Similarity 35.08; Pred. No. 6.4e-38;

DR PFAM; PF00684; DnaJ_CXXCXGXG; 1.

Query Match	29.2%	Score 549.5	DB 1	Length 418
Best Local Similarity	35.0%	Pred. No. 6.4e-38		
Matches 126; Conservative	75;	Mismatches 118;	Indels 41;	Gaps 10;

CC	ID	DNH_CUCSA	STANDARD;	PRT;	413 AA.
AC	004960;				
DT	01-FEB-1994 (Rel. 28, Created)				
DT	01-FEB-1994 (Rel. 28, Last sequence update)				
DT	01-NOV-1995 (Rel. 32, Last annotation update)				
DE	DNMJ PROTEIN HOMOLOG (DNMJ-1).				
GN	DNMJ1.				
OS	Cucumis sativus (Cucumber).				
OC	Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;				
OC	euphyllphytes: Spermatophyta; Magnoliophyta; eudicotyledons;				
OC	cory eudicots; Rosidae; eurosids 1; Cucurbitales; Cucurbitaceae;				
OC	Cucumis.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RP	TISSUE=SEED.				
RX	MEDLINE; 93343632.				
RA	Preisig-Mueller R., Kindl H.;				
RT	*Plant dnaJ homologue; molecular cloning, bacterial expression, and				
RT	expression analysis in tissues of cucumber seedlings.";				
RL	Arch. Biochem. Biophys. 305:30-37(1993).				
CC	-1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY				
CC	IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS.				
CC	-1- TISSUE SPECIFICITY: EXPRESSED IN SEEDLINGS IN ALL TISSUES, BUT				
CC	EXCEEDINGLY HIGH LEVELS IN HYPOCOTYLEDONS AND ROOTS.				
CC	-1- INDUCTION: WEAK, BY HEAT SHOCK				
CC	-1- SIMILARITY: CONTAINS A DNMJ-LIKE DOMAIN AND THE FOUR CXXCCGXG				
CC	MOTIFS FOUND IN DNMJ PROTEINS.				

Query Match	27.4%	Score 515.5	DB 1	Length 397
Best Local Similarity	34.5%	Pred. No. 3.7e-35		
Matches 133	Conservative 66	Mismatches 117	Indels 69	Gaps 13

```

Oy      26 FYKLIGVRSASIKDKKAVRKRLATOLPHRRDDPQAOEKPDIAGAVEVLSDESKRQ   85
Db      7 YDVLGVPRNMAQOEELKKAVRKRLAKLHYPKDPNNE--GEFKQTQSQAIEVLSDAKRREL   63
Oy      86 YDTYGESEELKDQH-----SSHGDIFSHFEGDFGFSGTTPROODNNIRGSIDIYDLERYT   141
Db      64 YDKGEQALKEGGACGGCGSPMDIFDMFG-----GGGRMRER--RGKNVWQLSTY   114
Oy      142 LEEVYAGNFVEEVNRKPV---ARQAPGKR--KC-NCR--QEMRTTGLGPGFFQMTO   189
Db      115 LEDLYNGATRKLAQLKNTVICDKCEGRGGKGGAVECCPMCRGTGMQIRIHQISPGWVOIQ   174
Oy      190 EYV-----CDECPNKLVNEERTLEVELEPEVCYRDGMEPFLEGEPHYDG   234
Db      175 SVCMEOGHGERISIPDRCKSCGNKRRIYREKKELLVHIDKGDOKITTFHEGEOBGL   234
Oy      235 EPGDLPFRFIAVKVHHPIFERRGDILTNTVSLVESLVGFEMDITHLDHKHVISR--DKI   292
Db      235 EPGDIIIVLDOKDAHVFTRRGEDLMCHMDIQVELACGFOKPISTLYDNRTIYTSHPOCI   294
Oy      293 TRPAAXMKKRGEGLEPNFNNNIKSILITTFDYDPFR-----EOLTEAREGIC   340
Db      295 VKHGDIKCVLNMGMVIYRRPEKGRLLIEFKNPENGPISPDKLSLEKLPERKE-VE   353
Oy      341 QLKQGVSQKV-----YNG   354
Db      354 ETDEMDOVELVDPEDPNERRRHYNG   378

RESULT      8
DNJ2_MOUSE  DNJ2_MOUSE STANDARD; PRT; 397 AA.
AC           P54102;
PT           01-OCT-1996 (Rel. 34, Created)
DT           01-OCT-1996 (Rel. 34, Last sequence update)
DT           15-JUL-1999 (Rel. 38, Last annotation update)
DE           DNAJ PROTEIN HOMOLOG 2 (HSJ-2).
GN           HSJ2 OR DNAJ2 OR RDJ1.
OS           Mus musculus (Mouse), and Rattus norvegicus (Rat).
OC           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
CC           Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN           [1]
RP           SEQUENCE FROM N.A.
RC           SPECIES-MOUSE;
RX           MEDLINE; 99017984.
RA           Royaux I., Minner F., Goffinet A.M., de Rouvroit C.L.;
RT           "A DnaJ-like gene, Hsj2, maps to mouse chromosome 5, at approximately
RL           24 cM from the centromere."
RL           Genomics 53:415-415(1998).
RN           [2]
RP           SEQUENCE FROM N.A.
RC           SPECIES-RAT; STRAIN-SPRAGUE-DAWLEY; TISSUE-BRAIN;
RA           Leng C.H., Wang C.;
RL           Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC           -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN AND THE FOUR CXXCXGXG
CC           MOTIFS FOUND IN DNAJ PROTEINS.
CC           -----
CC           This SWISS-PROT entry is copyright. It is produced through a collaboration
CC           between the Swiss Institute of Bioinformatics and the EMBL Outstation --
CC           the European Bioinformatics Institute. There are no restrictions in our way
CC           use by non-profit institutions as long as its content is in no way
CC           modified and this statement is not removed. Usage by and for commercial
CC           entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC           or send an email to license@sib.ch).
CC           -----
DR      EMBL; AF055664; AAC78597.1; -.
DR      EMBL; U53922; AAA98855.1; -.
DR      HSSP; P25685; 1HDJ.
DR      MGI; MGI:1270129; HSJ2.
DR      PRINTS; PR00625; DNAJPROTEIN.
DR      PROSITE; PS00636; DNAJ_1; 1.
DR      PROSITE; PS00637; DNAJ_CXXCXGXG; 1.
DR      PROSITE; PS50076; DNAJ_2; 1.

```

DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF00684; DnaJ_CXKXGXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 KW Chaperone; Repeat.
 FT DOMAIN 4 70 DNAJ-LIKE.
 FT DOMAIN 75 96 GLY-RICH.
 FT REPEAT 134 141 CXKXGXG MOTIF.
 FT REPEAT 150 157 CXKXGXG MOTIF.
 FT REPEAT 177 184 CXKXGXG MOTIF.
 FT REPEAT 193 200 CXKXGXG MOTIF.
 SQ SEQUENCE 397 AA: 44868 MW: 17836353D5C4CD558 CRC64;

Query Match 27.4%; Score 515.5; DB 1; Length 397;
 Best Local Similarity 34.5%; Pred. No. 3.7e-35;
 Matches 133; Conservative 66; Mismatches 117; Indels 69; Gaps 13;

QY 26 FYKILGVPRASIKDIKAYRKALQLHPDNDPPOAQEKFDLGAAYEVLSDSEKRRQ 85
 D 7 YIDVLGVPRNATQELKAYRKALKYHPDKNPNE--GEKFKQISQAYEVLADSKREL 63
 QY 86 YDTGEGEGLKDGHQ-----SSHGDFSHFGDFGFMFGSTPRQODRNIPRGSDIIVDEV 141
 D 64 YDKGGEQAIKSGAGGSGGSPMDFDFG-----GGGRMQRR--RGKNVHQLSVT 114
 QY 142 LEEVYAGNEVEVANKPV---ARQAPGR--KC--NCR--OEMRTTQGLGPRFQMTQ 189
 D 115 LEDLYNGATRKALQKNVICDKCEGREGKAGVCECPNCGTGMQIHQGPVMQIQ 174
 QY 190 EVV-----CDCEPVKLVNERTLEVELEGVGRGMEYPRIGSEEPVHG 234
 D 175 SVCEECQCHGRISPKRDCSCNGRKIVREKILVHIDGMKGQKITFGEDDQEPGL 234
 QY 235 EPGDLREIRIKVKKHIFERRGDDLYTNVTVSLVESLYGFEMDTHLGHVHISR--DKI 292
 D 235 EPGDIIVLDQKHAVFTREGEDLFMCMDIQLVALGCFQKPISTLNNRTIVITSHGQI 294
 QY 293 TRPGAXXMKKGEGLPNDNNIKSLIITFDVDFK-----EQLTEPARRGIK 340
 D 295 VKHEDICVLNEGPIYRREYKRLIIEFKVNPENGLSPDKLSLEKLPERKE--VE 353
 QY 341 QLLKQGSVKV-----YNG 354
 D 354 ETDEMDOVELVDPPNERRRRHNG 378

RESULT 9
 MAST_YEAST
 ID MAST_YEAST STANDARD; PRT; 409 AA.
 AC P25491;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MITOCHONDRIAL PROTEIN IMPORT PROTEIN MASS (PROTEIN YDJ1).
 GN MASS OR YDJ1 OR YNL064C OR N2418 OR YNL2418C.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 92107179.
 RA Atencio D.P., Yaffe M.P.;
 RT "MASS, a yeast homolog of DnaJ involved in mitochondrial protein
 RT import."
 RL Mol. Cell. Biol. 12:283-291(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE; 91332099.
 RA Caplan A.J., Douglas M.G.;
 RT "Characterization of YDJ1: a yeast homologue of the bacterial dnaJ
 RT protein."
 RL J. Cell Biol. 114:609-621(1991).
 RN [3]

RP SEQUENCE FROM N.A.
 RC STRAIN-S288C / FY1676;
 RX MEDLINE; 96021608.
 RA Berge P., Doignon F., Crouzet M.;
 RT "The sequence of a 44 420 bp fragment located on the left arm of
 RT chromosome XIV from Saccharomyces cerevisiae."
 RL Yeast 11:967-974(1995).
 RN [4]
 RP ERRATUM.
 RX MEDLINE; 97060022.
 RA Berge P., Doignon F., Crouzet M.;
 RL Yeast 12:297-297(1996).
 RN [5]
 RP ISOPRENOLD.
 RX MEDLINE; 92406811.
 RA Caplan A.J., Teal J., Casey P.J., Douglas M.G.;
 RT "Farnesylation of YDJ1p is required for function at elevated growth
 RT temperatures in Saccharomyces cerevisiae."
 RL J. Biol. Chem. 267:18890-18895(1992).
 CC -1- FUNCTION: PROBABLY INVOLVED IN MITOCHONDRIAL PROTEIN IMPORT. IS
 CC ALSO REQUIRED FOR EFFICIENT TRANSLLOCATION OF PRE-PRO-ALPHA-FACTOR.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED IN A PERINUCLEAR RING AS WELL
 CC AS IN THE CYTOPLASM (ACCORDING TO REF.2).
 CC -1- INDUCTION: MASS IS A HEAT SHOCK GENE WHOSE EXPRESSION INCREASES
 CC MODERATELY AT ELEVATED TEMPERATURES.
 CC -1- SIMILARITY: CONTAINS A DnaJ-LIKE DOMAIN AND THE FOUR CXKXGXG
 CC MOTIFS FOUND IN DnaJ PROTEINS.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; S74758; AAB20771.1; -
 DR EMBL; X56560; CAA39910.1; -
 DR EMBL; U12141; AAA99647.1; -
 DR EMBL; 271340; CAA95937.1; -
 DR PIR; A39659; A39659.
 DR PIR; S26703; S26703.
 DR HSSP; P25685; 1HDJ.
 DR SGD; L0002503; YDJ1.
 DR PRINTS; PS00625; DnaJPROTEIN.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PROSITE; PS00637; DnaJ_CXKXGXG; 1.
 DR PROSITE; PS00765; DnaJ_2; 1.
 DR PFAM; PF00226; DnaJ_1.
 DR PFAM; PF00684; DnaJ_CXKXGXG; 1.
 DR PFAM; PF01556; DnaJ_C; 1.
 KW Chaperone; Heat shock; Transport; Protein transport; Mitochondrion;
 KW Repeat; Prenylation; Lipoprotein.
 FT DOMAIN 4 72 DnaJ-LIKE.
 FT DOMAIN 75 103 GLY-RICH.
 FT REPEAT 143 150 CXKXGXG MOTIF.
 FT REPEAT 159 166 CXKXGXG MOTIF.
 FT REPEAT 185 192 CXKXGXG MOTIF.
 FT REPEAT 201 208 CXKXGXG MOTIF.
 FT LIPID 406 406 FARNESYL.
 SQ SEQUENCE 409 AA: 44670 MW: E4539F3618DD9CF2 CRC64;

Query Match 27.1%; Score 510; DB 1; Length 409;
 Best Local Similarity 36.6%; Pred. No. 1.1e-34;
 Matches 130; Conservative 56; Mismatches 119; Indels 50; Gaps 14;

QY 26 FYKILGVPRASIKDIKAYRKALQLHPDNDPPOAQEKFDLGAAYEVLSDSEKRRQ 85
 D 7 YIDVLGVPRNATQELKAYRKALKYHPDKNPSE--EAAEFKASAYEILSDPERDI 65
 QY 86 YDTGEGEGLKDGHSNS-----DIFSHFGDFGFMFGSTPRQODRNIPRGSDI 135
 D 115 LEDLYNGATRKALQKNVICDKCEGREGKAGVCECPNCGTGMQIHQGPVMQIQ 174

```

Db      66 YDQGEDEL-SGAGAGGPGGFGFDIPSGFFGAG-----AQRPGRQGRKID 117
Qy      136 VDLEVTLEEVYAGNFVEVYRKPV-ARQAPK-----RKC-NCRQE--MRTTQGP- 183
Db      118 HEISASLEELKYGRTAKLAKQILCKECEGRGKGAIVKCTSCNGGIFVTRQMPM 177
Qy      183 -GFRQ-----TOEVV-----CDECPNVKLVNEERTLEVEIPGVADMEYPIGGE 229
Db      178 IQRQTECDVCHGTGDIIDIPKDRCKSCNGKVKENERKLEHVHVGMDGQRIYFKGEAD 237
Qy      230 PHVDEPGLRFRKIVYKHPFERRGDLYTNTVSLVESLVGFEMDITHLDGH--KVHI 287
Db      238 QAPDVIGDVYFIVSEPRKSFKRDGDLVYEAFIDLTALAGEFALHSGMDLAKGI 297
Qy      288 SRDKITPRGAXXMKKSGEGLPNFDNNNKGSLITFPDVPKQQLTEARSGIKOL 342
Db      298 VPGVIAPGRMKVIEKGMP-IPRYGYGNLIFKTIKFPENHFTSE--ENMLKRL 349

RESULT 10
DNAB_BROUV STANDARD; PRT; 375 AA.
AC 005980;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DNAB PROTEIN.
GN DNAB.
OS Brucella ovis.
OC Bacteria: Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucella.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 25840 / 63/290T;
RX MEDLINE; 93094135.
RA Cellier M.F.M., Teyssier J., Nicolas M., Liautard J.P., Marti J.,
RA Sri Wadada J.;
RT "Cloning and characterization of the Brucella ovis heat shock protein
RT dnaJ functionally expressed in Escherichia coli.";
RL J. Bacteriol. 174:8036-8042(1992).
CC - FUNCTION: INTERACTS WITH DNAB, TO DISASSEMBLE A PROTEIN COMPLEX
CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
CC WITH GREP, THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).
CC - INDUCTION: BY HEAT SHOCK.
CC - SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC
CC DNAB-LIKE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M95799; AAC36133.1; -
DR EMBL; M94063; AAA93047.1; -
DR PIR; B47042; B47042.
DR HSSP; P08622; 1XBL.
DR PRINTS; PR00625; DNABPROTEIN.
DR PROSITE; PS00636; DNAB_1; 1.
DR PROSITE; PS50076; DNAB_2; 1.
DR PROSITE; PS00637; DNAB_CXXCXGXG; 1.
DR PFAM; PF00684; DnaJ_CXXCXGXG; 1.
DR PFAM; PF01556; DnaJ_C; 1.
KM Chapterone; DNA replication; Heat shock; Repeat.
FT DOMAIN 2 71 DNAB-LIKE.
FT DOMAIN 77 106 GLY-RICH.
FT REPEAT 146 153 CXXCXGXG MOTIF.
FT REPEAT 163 170 CXXCXGXG MOTIF.
FT REPEAT 185 192 CXXCXGXG MOTIF.
FT REPEAT 199 206 CXXCXGXG MOTIF.

```

```

SQ SEQUENCE 375 AA; 41021 MW; 61CB5F99FA2DC73D CRC64;
Query Match 24.9%; Score 469; DB 1; Length 375;
Best Local Similarity 32.8%; Pred. No. 2,36-31;
Matches 118; Conservative 59; Mismatches 129; Indels 54; Gaps 12;

Qy 25 DFKILGPPRSASIKDIKATRKALQLHPDRNPDPQAQKFPDGLGAAYLSSEK 84
Db 4 DYKALGVTATRADKTLKAAFRKLAHQYHPDRNPDPQAQKFPDGLGAAYLSSEK 63
Qy 85 QYDFYGEGLKDHQ-----SSHGDFSHFDFPFMGVGRPOODRNIPRSD 133
Db 64 AYDRFGHAFFERGGMGCGFGGFGAGGAFADIFEDFGE--MMGG--RRRRNGAERAD 118
Qy 134 IIVDLEVTLEEVYAGNFVEV-----YANKP-----VARQAPGRKRCNCRQEM 175
Db 119 LRYNMEVTLLEEVYAGNFVEVITICDESGSAGKPSQPTCTMGSGGRCRAQGF 178
Qy 176 -----RTTOLGGRFROMTOEVVCDCECPNVKLVNEERTLEVEI-----EPGYRDMETPFI 225
Db 179 FSVERTCPGCGRGQIILD-PCKKCHGCGRYDRGRSLSVNIPAVSRPEPRIR-----LA 231
Qy 226 GEGEPHVDEGP-GDLRFRKIVYKHPFERRGDLYTNTVSLVESLVGFEMDITHLDGK 284
Db 232 GEGEGLRGGRGADLYILSVKPHFFORDGADLYCKVPISTTTALGQEVSTLDGTQ 291
Qy 285 VHSRDKITPRGAXXMKKSGEGLPNFDNNNKGSLITFPDVPKQQLTEARSGIKOL 344
Db 292 TRVKTPECTQNGKQRLGKGMPL-RQSVTGDLTIQIDIEFP-QNLSKQRELEEFEX 349

RESULT 11
DNAB_ECOLI STANDARD; PRT; 375 AA.
AC P08622;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE DNAB PROTEIN.
GN DNAB OR GRP.
OS Escherichia coli.
OC Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-5.
RX MEDLINE; 8611849.
RA Ohki M., Tamura F., Nishimura S., Uchida H.;
RT "Nucleotide sequence of the Escherichia coli dnaJ gene and
RT purification of the gene product.";
RL J. Biol. Chem. 261:1778-1781(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 8611850.
RA Bardwell J.C.A., Tilly K., Craig E., King J., Zyliz M.,
RA Georgopoulos C.;
RT "The nucleotide sequence of the Escherichia coli K12 dnaJ gene. A
RT gene that encodes a heat shock protein.";
RL J. Biol. Chem. 261:1782-1785(1986).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE; 92334977.
RA Yura T., Mori H., Nagai H., Nagata T., Ishihama A., Fujita N.,
RA Isono K., Mizobuchi K., Nakata A.;
RT "Systematic sequencing of the Escherichia coli genome: analysis of
RT the 0-2.4 min region.";
RL Nucleic Acids Res. 20:3305-3308(1992).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE; 97426617.

```

ID	NAME	STANDARD	PRT	378 AA
DB	61 KRAAVDYGNAHAFQGGGMMGGGGGREGGAGDESDIDGVDFGDIFFGGGGRGQ--RAARGADLRL			118
QY	137 DLEVTLEEVYAGNEFVEVVRNKPVARQA-----PGRKRCNCRQEMKTTOLQGPGRFM			187
DB	119 NMELTLEAVAGVTKET--RIPTLEEDCYCHSGSAGKPTGPOTC-----PTCHSGGVQM			171
QY	168 TQ-----EWCDECPVVKVYNRETFLEIEIPRGVSDCKEVPFTIGEG			228
DB	172 RQGFPAVQTCPRHCGGGLTKIPRCCKCHGHGVEBSKTLSTKIPAGVDTDRIRLAGEG			231
QY	229 RPHVDGEF-GDLRRIRIVVKNPIFEERRGDDLTNTVTSLVESVGFENDITHLDGKHVHI			287
DB	232 IAGEGAPAGDLVQVQVQKHQPIFEERGNNTLCSEVINFPAALAGEIIEVPTLDG-RVKL			290
QY	268 SRDKITPGAXXMKKSGEGLEPNFNNNIKSLLIITDYDFP-----KQLTLEARE			337
DB	291 KVPGETQIGKIFRRKRGKVASV-RGSAQGDLLCRVAVETPVGLANEROKOLLQELQE			345
RESULT	12			
DNAL_SALTY				
ID	DNAL_SALTY	STANDARD	PRT	378 AA
AC	060004			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	15-DEC-1998 (Rel. 37, Last annotation update)			
DE	DNAL PROTEIN.			
GN	DNAL.			
OS	Salmonella typhimurium.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
CC	Salmonella.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-LT2;			
RA	Stephen R.J., Hinton J.C.D.;			
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.			
CC	-I- FUNCTION: INTERACTS WITH DNAL, TO DISASSEMBLE A PROTEIN COMPLEX			
CC	AT THE PHASE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY			
CC	WITH GREP, THE ATPASE ACTIVITY OF DNAL.			
CC	-I- SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC	-I- INDUCTION: BY HEAT SHOCK UNDER THE CONTROL OF THE HTPR REGULATORY			
CC	PROTEIN (BY SIMILARITY).			
CC	-I- SIMILARITY: TO OTHER PROKARYOTIC DNAL, AND TO EUKARYOTIC			
CC	DNAL-LIKE PROTEINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/lsb.ch).			
CC	or send an email to license@lsb-sib.ch .			
CC	-----			
DR	EMBL; U58360; AAB02911.1; -.			
DR	HSSP; P08622; 1XBL.			
DR	STRGENE; SG27272?; DNAL.			
DR	PRINTS; PR00625; DNALPROTEIN.			
DR	PROSITE; PS00636; DNAL_1; 1.			
DR	PROSITE; PS50076; DNAL_2; 1.			
DR	PROSITE; PS00637; DNAL_CXXCXGXG; 1.			
DR	PFAM; PF00226; Dnal; 1.			
DR	PFAM; PF00684; Dnal_CXXCXGXG; 1.			
DR	PFAM; PF01556; Dnal_C; 1.			
KW	Chaperone; DNA replication; Heat shock; Repeat.			
FT	INIT_MET	0		
FT	DOMAIN	2	71	
FT	DOMAIN	76	112	
FT	DOMAIN	146	153	
FT	REPEAT	163	170	
FT	REPEAT	185	192	
FT	REPEAT	199	206	
FT	SEQUENCE	378 AA;	41181 MW;	176ECC64696F30A2 CRC64;

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AF007813; AAC35417.1; -
 DR EMBL: U72647; AAB17396.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DNAS_1; 1.
 DR PROSITE: PS00076; DNAS_2; 1.
 DR PROSITE: PS00637; DNAS_CXXCXGXG; FALSE_NEG.
 DR PFAM: PF00226; Dnas; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 3 72
 FT REPEAT 150 157 CXXCXGXG MOTIF.
 FT REPEAT 167 174 CXXCXGXG MOTIF.
 FT REPEAT 189 196 CXXCXGXG MOTIF.
 FT REPEAT 203 210 CXXCXGXG MOTIF.
 SO SEQUENCE 369 AA; 39939 MW; 408B5B11457DCDF CRC64;

Query Match 23.4%; Score 441.5; DB 1; Length 369;
 Best Local Similarity 31.0%; Pred. No. 4.1e-29;
 Matches 117; Conservative 64; Mismatches 135; Indels 61; Gaps 12;

QY 24 RDYKYLIGVRSASINDIKAYRKALQLHPDNPDPQAEKFDLGAAYEVLSDSEKR 83
 DB 4 RSTYDLIGVRSKANDERIKSAVYRLAKYHPDNKNGKESEKFEKATEYETILRDPKRR 63
 QY 84 KOYDTYGEELK-----DGHOSHGDIFSHFGDF-----GFMEGTPRODDRNIPRG 131
 DB 64 QAVDQEGKAGVSGAGGFGGATDPDIDIGDIFGDFGSGRSGSGGGRSPQ---RG 120
 QY 132 SDIIVDLVLELVYVAGN--FVEVVR-----NKPYARQAPGRKRC-NCRQEMRTTQLG 181
 DB 121 SDRYVNLVSLIEDNALGREGYKIEIPRLSECVDCNGSGASGSSPACPCDCG-----G 172
 QY 182 PGRFOMTQEVY-----CDGCPNVKLVNEERTLEVELEPGRVRODMEX 222
 DB 173 SGQIRRTQGGFFSVATTCPTCRGKGTIISNPNCRSGGGLQEKRTTINIKIPGVEYTSRL 232
 QY 223 PRTGEGEPHYGEP-GDLRFRIKVVHPIFERRGDLYTNVTVSLVESLVGFMDITTHLD 281
 DB 233 KVSQEGEAGNGSPHGDLYVVTIKKHELEFRGNDLIVRKISLGAIGALIEVPTID 292
 QY 282 GKHVHISRDKITRPGAXXMKKGGELPNFDNNNIKSLITFDVDFPREQLTEPAREGIQ 341
 DB 293 GKAKKKIKEGTESGQVFLRKGHMPYLGAYG-KGDQHVIVKIEIPK-KITRRQRELIEE 350
 QY 342 LKQ-----GSYQVY 352
 DB 351 FAREGENTIPGSKITF 367

RESULT 15
 DNAS METTH STANDARD: PRT; 376 AA.
 AC 027352;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE DNAS PROTEIN.
 GN DNAS OR MTH1291.
 OS Methanobacterium thermoautotrophicum.
 OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
 OC Methanobacterium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-DELTA H:
 RK MEDLINE; 98037514.

RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
 RA Haddage T., Bashirzadeh R., Blakey D., Cook R., Gilbert K.,
 RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
 RA Spadofora R., Viscare R., Wang Y., Wierzbowski J., Gibson R.,
 RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
 RA McDougall S., Shiner G., Goyal A., Pietrowski S., Church G.M.,
 RA Daniels C.J., Mao J.-I., Rice P., Nollan J., Reeve J.N.,
 RT "Complete genome sequence of Methanobacterium thermoautotrophicum
 RT deltaH: functional analysis and comparative genomics.",
 RU J. Bacteriol. 179:7135-7153(1997).
 CC -1- FUNCTION: STIMULATES, JOINTLY WITH GRPE, THE APPASE ACTIVITY OF
 CC DNAS (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAS, AND TO EUKARYOTIC
 CC DNAS-LIKE PROTEINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: AE000894; AAB85773.1; -
 DR HSSP: P08622; 1XBL.
 DR PRINTS: PR00625; DNASPROTEIN.
 DR PROSITE: PS00636; DNAS_1; 1.
 DR PROSITE: PS00076; DNAS_2; 1.
 DR PROSITE: PS00637; DNAS_CXXCXGXG; FALSE_NEG.
 DR PFAM: PF00226; Dnas; 1.
 DR PFAM: PF00684; Dnas_CXXCXGXG; 1.
 DR PFAM: PF01556; Dnas; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 2 71
 FT REPEAT 148 155 CXXCXGXG MOTIF.
 FT REPEAT 165 172 CXXCXGXG MOTIF.
 FT REPEAT 191 198 CXXCXGXG MOTIF.
 FT REPEAT 205 212 CXXCXGXG MOTIF.
 SO SEQUENCE 376 AA; 41584 MW; 3497C7DA61DC5342 CRC64;

Query Match 23.4%; Score 441; DB 1; Length 376;
 Best Local Similarity 34.3%; Pred. No. 4.6e-29;
 Matches 126; Conservative 55; Mismatches 128; Indels 58; Gaps 14;

QY 21 IAGRDYKILGVRSASINDIKAYRKALQLHPDNPDPQAEKFDLGAAYEVLSDS 80
 DB 1 MARQDYELIGVDRGADKKIKAYRRLARKYHPDVS-DDPAAEKFEKISRAYAVLSDD 59
 QY 81 ERKQYDTYGEELKNG-----HOSHGDIFSHFGDF-----FMGSGTPRODDRN 128
 DB 60 EKRARDYRGHAGM-DGSEQEDIFNNINFEIDIFSGLGFIDIGLFGGGRRRHGPQ--- 117
 QY 129 PRGSDIIVDLVLELVYVAGNVEV---VRNKPV---ARQAPGRKKNCR-----Q 173
 DB 117 -RGADISTYLDISLSDAYNGLETDIVRPTKPCPYCHGSRABGRTGTCQCGSGGQVR 175
 QY 174 EMRTTOLGPRGQMTQ-----EYVCDECPNVKLVNEERTLEVELEPGRVQDG 219
 DB 176 QVKNITLQ---QMNNTTCCPDGCGEGYVEKPCSNCKNGGVVRRKSTIHVVPAGVETG 231
 QY 220 MEYPTGEGEPHY-DGEPGDLRFRIKVVHPIFERRGDLYTNVTVSLVESLVGFMDIT 278
 DB 232 SRLRVPGESEMGIRGEPDLYVVIKVKHSTFRREGANLVTEKPTISYQALGLTVEVP 291
 QY 279 HLDGKHVHISRDKITRPGAXXMKKGGELPNFDNNNIKSLITFDVDFPREQLTEPAREG 338
 DB 292 TLD-RPVKLRIPAGTQSGTFVFKVKGHMPHLMKNGY-GNLVYKVVVPRKLSPPQ---- 346
 QY 339 IKQLAKQ 345
 DB 346 -KELLRE 351

Sat Jul 1 21:34:08 2000

us-09-501-714-1.rsp

Page 11

Search completed: June 30, 2000, 11:56:22
Job time: 7137 sec

THIS PAGE BLANK (USPTO)

Db 307 VKPDPQFAINDEGMPMYQRPFMKGKLYIHFSVPEPDSLNEQCKALGGV--LPPRTSVQ 363

RESULT 5
T07371
dnaJ protein homolog - potato
C:Species: Solanum tuberosum (potato)
C:Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07371
R:Legewie, G.; Braun, H.P.
Plant Physiol. 117, 1127, 1998
A:Title: A cDNA from potato with homology to DnaJ is identical to a hitherto unidentified
A:Reference number: Z15992
A:Accession: T07371
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-419 <LEG>
A:Cross-references: EMBL:X94301; NID:g1125690; PIDN:CAA63965.1; PID:e214633
A:Experimental source: tissue type leaf
C:Genetics:
A:Gene: dnaJ
A:Note: Induced upon tuberization
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
C:Keywords: molecular chaperone

Query Match 28.4%; Score 534; DB 2; Length 419;
Best Local Similarity 36.7%; Pred. No. 4,8e-37;
Matches 130; Conservative 68; Mismatches 112; Indels 44; Gaps 13;

QY 26 FYKILGVPKRSIKDKAYRKALQLPDRNPDPQAEKFDLGAAYVLSDESKRKQ 85
DB 14 YETILGVPKTAAGBDLKAAYRKALKNHPPKGGD---PEKFKLAQAYELSDPERKEI 69
QY 86 YDTYGEGLKDGHQSSHG-----DIFSHFFG--DFGFMFGTPROODRNIIPGSDIYDL 138
DB 70 YDQYGEALKEKMGGGGGGMDPFDIFSSFFGSGPFG---GGGSSRRQRGRGDEVVHPL 126
QY 139 EYTLLEVYAGNF--VEVYRN-----KVARQAPGRKCN-CR-QEMRTT--QLGPRFQ 186
DB 127 KVSLEDLYNGTSKSLSLSRNVLCSCCKGSKSGASMKSCGCGSGMKYITRLQGPMSIQ 186
QY 187 MTQ-----EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFGESEPH 231
DB 187 QMHPNCECKGTGEMINDKRCGQCKGKGVQEKVLEVVVEKMGONGQKITFFGEADEA 246
QY 232 VDEPGLDPRIRIKVYKHPFERRGGDLTYNTVYSLVESLVGFEMDITHLDGKHVHI--SR 289
DB 247 PDVTGDIYVLDQKEHPKRRKGGDLFEVHTLSLTALCGFQFIITHLDNRQLITKPA 306
QY 290 DKITRPGAXXMKKGEGLPNFDNNNIKSLITFDVDPKQDLTEAREGIKOLL 343
DB 307 GEVYKRPQFAINDEGMPMYQRPFMKGKLYIHFTVEFP-DTLSPQCKNLEAVL 359

RESULT 6
JQ2142
chaperone ANO1 protein - Atriplex nummularia
C:Species: Atriplex nummularia
C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 05-Dec-1997
C:Accession: JQ2142
R:Zhu, J.K.; Shi, J.; Bressan, R.A.; Hasegawa, P.M.
Plant Cell 5, 341-349, 1993
A:Title: Expression of an Atriplex nummularia gene encoding a protein homologous to the
A:Reference number: JQ2142; MUID:93222693
A:Accession: JQ2142
A:Molecule type: mRNA
A:Residues: 1-417 <ZHU>
A:Cross-references: GB:I.09124
C:Comment: This protein is involved in protein folding and is observed as heat shock pr
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
C:Keywords: molecular chaperone

F:13-74/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 28.0%; Score 526.5; DB 2; Length 417;
Best Local Similarity 34.9%; Pred. No. 2e-36;
Matches 123; Conservative 70; Mismatches 118; Indels 41; Gaps 11;

QY 26 FYKILGVPKRSIKDKAYRKALQLPDRNPDPQAEKFDLGAAYVLSDESKRKQ 85
DB 14 YETILGVPKTAAGBDLKAAYRKALKNHPPKGGD---PEKFKLAQAYELSDPERKEI 69
QY 86 YDTYGEGLKDGHQSSHG-----DIFSHFFGDFGFMFGTPROODRNIIPGSDIYDLE 140
DB 70 YDQYGEALKEKMGGGGGMDPFDIFSSFFGSGPFG---GSPGGSSRRQRGRGDEVVHPLK 127
QY 141 TLEEVYAGNFVEV--VRKPVARQAPGR-----RKCN-CR--QEMRTTQLGPRPMT 188
DB 128 SLEDLFTGTTRKSLSLSRNVLCSCCKGSKSGASMKSCGCGTGMKYSIRHLGPMISIQ 187
QY 189 Q-----EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFGESEPH 233
DB 188 QHPNCECKGTGETINDKRCPCQCKGKGVQEKVLEVVVEKMGONGQKITFFGEADAPD 247
QY 234 GEPGLDPRIRIKVYKHPFERRGGDLTYNTVYSLVESLVGFEMDITHLDGKHVHSRD--K 291
DB 248 TVTSDIYFVLDQKEHPKRRKGGDLFEVHTLSLTALCGFQFIITHLDNRQLITKSNLGE 307
QY 292 ITRPGAXXMKKGEGLPNFDNNNIKSLITFDVDPKQDLTEAREGIKOLL 343
DB 308 VKPDPQFAINDEGMPMYQRPFMKGKLYIHFTVEFP-DLNPDPQVLSLEAIL 358

RESULT 7
S33312
dnaJ protein - leek (fragment)
C:Species: Allium porrum (leek)
C:Date: 06-Jan-1995 #sequence_revision 01-Sep-1995 #text_change 20-Aug-1999
C:Accession: S33312; S31387
R:Bessoule, J.J.
FEBS Lett. 323, 51-54, 1993
A:Title: Occurrence and sequence of a DnaJ protein in plant (Allium porrum) epidermal
A:Reference number: S33312; MUID:93265942
A:Accession: S33312
A:Molecule type: mRNA
A:Residues: 1-397 <BES>
A:Cross-references: EMBL:X69436; NID:g16086; PIDN:CAA49211.1; PID:g16087
C:Genetics:
A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
F:1-53/Domain: dnaJ amino-terminal homology (fragment) <DNJ>

Query Match 27.6%; Score 519.5; DB 2; Length 397;
Best Local Similarity 35.6%; Pred. No. 7.2e-36;
Matches 124; Conservative 66; Mismatches 117; Indels 41; Gaps 9;

QY 34 RSASIKDKIKAYRKALQLPDRNPDPQAEKFDLGAAYVLSDESKRKQYTYEEG 93
DB 1 KNASPDLLKKAAYRKALKNHPPKGGD---PEKFKLAQAYELSDPERKEIYDQYEDA 56
QY 94 LKDGHQSSHG-----DIFSHFFGDFGFMFGTPROODRNIIPGSDIYDLEVTLEEVYAG 148
DB 57 LKEGMGGGGGMDPFDIFSSFFGSGGSSR--GRQRRGDEVVHPLKSLLELYNG 114
QY 149 NFVYVANKRPV-----AQAPGRKKNCR--QEMRTTQLGPRPMTQ----- 190
DB 115 TSKLISLRNVYICSKGKSGASRCASGCGSGMKVSLRGLGPMIQMGHPNDCK 174
QY 190 ---EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMETPFGESEPHVDEGPDRL 241
DB 175 GTGEMINDKRCPLCKGKGVQEKVLEVVVEKMGONGQKITFFGEADAPDVTGDIYF 234
QY 242 RIKVYKHPFERRGGDLTYNTVYSLVESLVGFEMDITHLDGKHVHSRD--KITRPGAXX 299

Db 246 HVLPGMKHNDKITFKGDGDQSDPDGPDGVIVIQKDHDIKRDGDDLHMTKRLSNEA 305
 QY 270 LVGFEMDITLHGHKVNHS--RDKITRGAAXXMKKGEGELPFPDNNNIGSLITFDVDP 327
 Db 306 LCGYFELIKHLDGHPVLVSSKQGVYIKGVIRGVLGKMPMKKYPKELGNLFEEVEFP 365
 QY 328 KEQLETFAR 336
 Db 366 KEHFLDDEK 374

RESULT 13

E70361
 A:Chaperone DnaJ - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 20-Aug-1999
 C:Accession: E70361
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; O'V.
 Nature 392, 353-358, 1998
 A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:98196666
 A:Accession: E70361
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-376 <AOF>
 A:Cross-references: GB:AE000703; NID:92983287; PIDN:AN06881.1; PID:92983289; GB:AE00065
 A:Experimental source: strain VFS
 C:Genetics:
 A:Gene: dnaJ2
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:8-72/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 26.0%; Score 490; DB 2; Length 376;
 Best Local Similarity 34.5%; Pred. No.2e-33;

Matches 128; Conservative 61; Mismatches 118; Indels 64; Gaps 15;

QY 24 RDEYKILGVPRASIKDKKAYRKALQLHPDRNPDDPOAERFODLGAAYVLSDESK 83
 Db 7 KDYEELIGVPRNASQOEIKKAYRLVAKHPD-ICKPECEKEKKEINAYVLSDESK 65
 QY 84 KQDVTYEEELKNGHSHG-----DISHFEG-DEGFNF---GGTPRO 123
 Db 66 KLYDMT-----GHAFFEGAGAOQVETTELPIREILREFFDIFGIFERATGRRRA 118
 QY 124 QDRNIPRGSIIIVLEVTLEEVYAGNFVEVVRNKPVARQAPG-----KRC-NCRQE 174
 Db 119 RRRRSVKGEDIVVPEITTELEAFKGTTPLEVEREVPSCAGGTGIDESKRTCPGGR 178
 QY 175 MRTLQGPGRFQMTQEVVDCDPNKLVE-----ERTLEVEIEPGRDGM 221
 Db 179 GETVQ-GNMFQYRQ--TCPTCGEGVYENCHACTGRGYGLVKELIKVIRPPGVADGR 235
 QY 222 YPIGGEPEV---DDEPGLRIRIVVKKPIFERGGDLITVWYVLSVLFEMDIT 278
 Db 236 --LVVEGKGAGAYGPGDLIYIVVKKPIFERGGDLITVWYVLSVLFEMDIT 293
 QY 279 HLDGKHVHSRDKITRPGAXXMKKGELPFPDNNNIGSLITFDVDPK---EQLTSE 334
 Db 294 TLDEKVVKKIRPPCTKEGELIKVPGKMPRLKSG-RGDLIVVNHIDYKIGVLSKLTD 352
 QY 335 AREGILQLKQ 345
 Db 353 GRK-VEELKQ 362

RESULT 14

T22648
 A:Hypothetical protein F54D5.8 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T22648

R:Coles, L.; Matthews, L.
 submitted to the EMBL data library, October 1995
 A:Reference number: Z19593
 A:Accession: T22648
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-331 <NTL>
 A:Cross-references: EMBL:266513; NID:e1062016; PIDN:CA91334.1; GSPDB:GN00020; CESP:F
 A:Experimental source: clone F54D5
 C:Genetics:
 A:Gene: CESP:F54D5.8
 A:Map position: 2
 A:Introns: 46/3; 289/2

Query Match 25.0%; Score 470.5; DB 2; Length 331;
 Best Local Similarity 32.4%; Pred. No.7.1e-32;

Matches 119; Conservative 55; Mismatches 102; Indels 91; Gaps 9;

QY 23 GRDPEYKILGVPRASIKDKKAYRKALQLHPDRNPDDPOAERFODLGAAYVLSDESK 82
 Db 2 GKDYIKVLGISKATDDEIKKAYRKALKYHPDKN-KEGAENKFEKIEAAYDVLSDDPK 60
 QY 83 RKQDVTYEEELKNGD-----HQSSHD---IFSHFGD----- 113
 Db 61 KKIYDQFEGEGELKEGGAGGGGGGGMHYEFRRDPNNITSSFFGSGDPPGAGGPFIDG 120
 QY 113 -----FGFMFGTPRQDDRNIPRGSIIIVLEVTLEEVYAGNFVE 152
 Db 121 GGAGGPFMEFMNOGMDNDFGGMHOGGRGHARQPA---VLHDLVSLEVLKNG---- 174
 QY 153 VVNRKPVARQAPGRKRCNCRQEMRTQLGPRPOMQEVVDCDPNKLVEERTLEVEI 212
 Db 174 -----TTK-----KKITRKVINDAORL---EDKVLVTI 201
 QY 213 EPGVRDGMETPIEGEPHYDGPDLRPIRVKVPKPIFERGGDLITVWYVLSVESTIG 272
 Db 202 KPGMKSGTKITPFKEGDQHPNRPADIYVYKDKPPIKPKRGGSDIKRVEKISLSALIG 261
 QY 273 FEVDIHLHGHKVNHSRDKITRPGAXXMKKGELPFPDNNNIGSLITFDVDPREQLT 332
 Db 262 LDIMITLGGADYRLQNLNVIRKPTRRLTGLPMPKPSHRGDLITFEVDFP-SQLN 320
 QY 333 EEARREGI 339
 Db 321 PTOGREVI 327

RESULT 15

B47042
 A:heat shock protein dnaJ - Brucella ovis

C:Species: Brucella ovis
 C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 20-Aug-1999

C:Accession: B47042
 R:Cellier, M.F.; Teyssier, J.; Nicolas, M.; Liautard, J.P.; Marti, J.; Sri Widadada, J.
 J. Bacteriol. 174, 8036-8042, 1992

A:Title: Cloning and characterization of the Brucella ovis heat shock protein dnaJ fu
 A:Reference number: A47042; MUID:93094135

A:Accession: B47042
 A:Status: preliminary

A:Molecule type: nucleic acid
 A:Residues: 1-375 <CEL>

A:Cross-references: EMBL:M95799; NID:9144114; PIDN:AN06133.1; PID:9144116
 A>Note: sequence extracted from NCBI backbone (NCBIN:119966, NCBI:P:119968)

C:Genetics:
 A:Gene: dnaJ

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

C:Keywords: DNA replication; heat shock; molecular chaperone

F:4-69/Domain: dnaJ amino-terminal homology <DNJ>
 F:76-108/Region: G/F motif
 F:146-153/Region: CXXCXGXG repeat
 F:163-170/Region: CXXCXGXG repeat
 F:185-192/Region: CXXCXGXG repeat

F:199-206/Region: CXXCXGXG repeat

```
Query Match      24.9%; Score 469; DB 2; Length 375;
Best Local Similarity 32.8%; Pred. No. 1.1e-31;
Matches 118; Conservative 59; Mismatches 129; Indels 54; Gaps 12;

QY 25 DEYKIIIVPRASIKDKKAYRKALQLHPDRPDPOAQEKFODLGAAYEVLSDSERK 84
   ||: ||| ||: ||: ||: ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 DYEAALGVTRTADDKTLKAAFKRLAQYHPDRPDDEAEERKFEIGEAETLKDPOKRA 63
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY 85 QYDYGEEGELKDGHQ-----SSHGDIFSHFEGDFGFMFGGTPROODRNIPRGS 133
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 64 AYDRFGHAAFEENGGMGGFNGFAGAGGFADIFEDIFGE---MMCG--RRRRGAERKAD 118
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY 134 IIVDLEVTLEEVYAGNFEVY-----VRNKP-----VARQAPGKRKCNCROEM 175
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 119 LRYNMEVTLLEAAYAGKTAQIRVPTSLTCDSCSGSAGAPGSQPTTCTWCSGSGCRAAQGF 178
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY 176 ----RTQLGPRGRQMTQEVVDCENVKLVNEERTLEVEI-----EPGVROGMEYPT 225
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 179 FSYERTCPGCGNGRGOI IKD--PCKKCHGQGVRDGRSLSVNIPAVSRTEPRIR-----LA 231
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY 226 GEGEPIVYDEEP-GDLRPRIKIVKHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGK 284
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 232 GESEAGLRGSPAGDLYLLSVKPFHEFFQKDGADLYCKVPISMTTAALGQGFVSTLDGTQ 291
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
QY 285 VHSIRDKITRPGAXXWKKGEGLPNFDNNNIKSLIITFDVDFPKEQLTPEARBGIKQLK 344
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 292 TRYKVEPGTQNGKQFRLKKGMPGL-RQSVTGDLYIQIDIEIP-QNLSKRQRELLSEFEK 349
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
```

Search completed: June 30, 2000, 09:58:19
Job time: 939 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:13:20 ; Search time 16.52 Seconds
(without alignments)
312.869 Million cell updates/sec

Title: US-09-501-714-1

Perfect score: 1883

Sequence: 1 MAPONLSTFCLLLYLIGAV.....IKOLKQSGSVKYNGLQGY 358

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep:*
4: /cgn2_6/prodata/2/1aa/PCRTUS_COMB.pep:*
5: /cgn2_6/prodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1879	99.8	358	2	US-08-868-288A-1	Sequence 1, Appl1
2	1879	99.8	358	3	US-09-235-373-1	Sequence 1, Appl1
3	1879	99.8	358	3	US-09-388-993-1	Sequence 1, Appl1
4	515.5	27.4	397	2	US-08-868-288A-5	Sequence 5, Appl1
5	515.5	27.4	397	3	US-09-235-373-5	Sequence 5, Appl1
6	515.5	27.4	397	3	US-09-388-993-5	Sequence 5, Appl1
7	515	27.3	419	2	US-08-868-417-3	Sequence 3, Appl1
8	448.5	23.8	348	2	US-08-868-417-1	Sequence 1, Appl1
9	437	23.2	340	2	US-08-974-546-5	Sequence 1, Appl1
10	345	18.3	350	2	US-08-472-534-6	Sequence 6, Appl1
11	257	13.6	330	2	US-08-868-288A-3	Sequence 3, Appl1
12	257	13.6	330	2	US-09-235-373-3	Sequence 3, Appl1
13	257	13.6	330	3	US-09-388-993-3	Sequence 3, Appl1
14	231.5	12.3	277	2	US-08-868-288A-7	Sequence 7, Appl1
15	231.5	12.3	277	3	US-09-235-373-7	Sequence 7, Appl1
16	231.5	12.3	277	3	US-09-388-993-7	Sequence 7, Appl1
17	231.5	12.3	351	2	US-08-868-288A-6	Sequence 6, Appl1
18	231.5	12.3	351	3	US-09-235-373-6	Sequence 6, Appl1
19	231.5	12.3	351	3	US-09-388-993-6	Sequence 6, Appl1
20	194	10.3	438	2	US-08-897-340-34	Sequence 34, Appl1
21	192	10.2	484	2	US-08-879-260-4	Sequence 4, Appl1
22	154.5	8.2	51	2	US-08-346-849-12	Sequence 12, Appl1
23	154.5	8.2	51	2	US-08-293-284A-8	Sequence 8, Appl1
24	154	8.2	52	2	US-08-293-284A-8	Sequence 8, Appl1
25	154	8.2	52	2	US-08-346-849-10	Sequence 10, Appl1
26	151	8.0	52	1	US-08-293-284A-10	Sequence 10, Appl1
27	151	8.0	52	1	US-08-346-849-9	Sequence 9, Appl1
28	145	7.7	52	1	US-08-293-284A-9	Sequence 9, Appl1
29	145	7.7	52	2	US-08-293-284A-9	Sequence 9, Appl1

30	121.5	6.5	49	1	US-08-346-849-11	Sequence 11, Appl1
31	121.5	6.5	49	2	US-08-293-284A-11	Sequence 11, Appl1
32	100.5	5.3	550	2	US-08-844-058-2	Sequence 2, Appl1
33	99	5.3	433	1	US-08-346-849-2	Sequence 2, Appl1
34	99	5.3	433	2	US-08-293-284A-2	Sequence 2, Appl1
35	97.5	5.2	51	1	US-08-346-849-14	Sequence 14, Appl1
36	97.5	5.2	51	2	US-08-293-284A-14	Sequence 14, Appl1
37	95	5.0	56	1	US-08-346-849-28	Sequence 28, Appl1
38	95	5.0	56	2	US-08-293-284A-28	Sequence 28, Appl1
39	89	4.7	3135	1	US-08-323-1708-2	Sequence 2, Appl1
40	86	4.6	459	2	US-08-481-6588-2	Sequence 2, Appl1
41	86	4.6	459	2	US-08-477-504A-2	Sequence 2, Appl1
42	86	4.6	459	2	US-08-486-756A-2	Sequence 2, Appl1
43	86	4.6	459	2	US-08-485-8628-2	Sequence 2, Appl1
44	86	4.6	459	3	US-08-787-739-2	Sequence 2, Appl1
45	86	4.6	1482	1	US-08-026-138E-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-868-288A-1
; Sequence 1, Application US/08868288A
; Patent No. 5922567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preethi
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Inocyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fastseq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,288A
; FILING DATE: June 3, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-08-868-288A-1

Query Match 99.8%; Score 1879; DB 2; Length 358;
Best Local Similarity 100.0%; Pred. No. 2.3e-199;
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Gy 1 MAPONLSTFCLLLYLIGAVAGADFYKILGVPSASIKDKKAYRLALQLHPDRPPD 60

```
Db 1 MAPQNSTCCLLLLYLIGAVIAGRDYKILGVPRASISINDIKAYRKLAQLHPDRNPDD 60
Qy 61 POAEKFQDLGAAYEVLVSSEKRRKQDYTYGEEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Db 61 POAEKFQDLGAAYEVLVSSEKRRKQDYTYGEEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Qy 121 PROQDNRIPRGSIDIYDLEVTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTTL 180
Db 121 PROQDNRIPRGSIDIYDLEVTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTTL 180
Qy 181 GGRFQMTQEVVCDCEPNKLVNEERTLEVEIEPGVRDMEYPFIEGEPHVDGEPDGLR 240
Db 181 GGRFQMTQEVVCDCEPNKLVNEERTLEVEIEPGVRDMEYPFIEGEPHVDGEPDGLR 240
Qy 241 FRKVVYKHPFERRGDLYTNVTVSLVESLVGFEMDITLHDGKHVHISRDKITRPGAXXW 300
Db 241 FRKVVYKHPFERRGDLYTNVTVSLVESLVGFEMDITLHDGKHVHISRDKITRPGAXXW 300
Qy 301 KKEGGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIGIKOLLKQGSYQKVYNGLOGY 358
Db 301 KKEGGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIGIKOLLKQGSYQKVYNGLOGY 358
```

RESULT 2

```
US-09-235-373-1
; Sequence 1, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-235-373-1
```

Query Match 99.8%; Score 1879; DB 3; Length 358;
Best local Similarity 100.0%; Pred. No. 2.3e-199;

```
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MAPQNSTCCLLLLYLIGAVIAGRDYKILGVPRASISINDIKAYRKLAQLHPDRNPDD 60
Db 1 MAPQNSTCCLLLLYLIGAVIAGRDYKILGVPRASISINDIKAYRKLAQLHPDRNPDD 60
Qy 61 POAEKFQDLGAAYEVLVSSEKRRKQDYTYGEEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Db 61 POAEKFQDLGAAYEVLVSSEKRRKQDYTYGEEGLKDGHOSSHGDISSHFGDFGFMFGGT 120
Qy 121 PROQDNRIPRGSIDIYDLEVTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTTL 180
Db 121 PROQDNRIPRGSIDIYDLEVTLEEVYAGNFVEVVRNKPVARQAPGRKNCROEMKTTL 180
Qy 181 GGRFQMTQEVVCDCEPNKLVNEERTLEVEIEPGVRDMEYPFIEGEPHVDGEPDGLR 240
Db 181 GGRFQMTQEVVCDCEPNKLVNEERTLEVEIEPGVRDMEYPFIEGEPHVDGEPDGLR 240
Qy 241 FRKVVYKHPFERRGDLYTNVTVSLVESLVGFEMDITLHDGKHVHISRDKITRPGAXXW 300
Db 241 FRKVVYKHPFERRGDLYTNVTVSLVESLVGFEMDITLHDGKHVHISRDKITRPGAXXW 300
Qy 301 KKEGGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIGIKOLLKQGSYQKVYNGLOGY 358
Db 301 KKEGGLPNFDNNNIKSLIITFDVDPKQQLTEAREGIGIKOLLKQGSYQKVYNGLOGY 358
```

RESULT 3

```
US-09-388-993-1
; Sequence 1, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 358 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-388-993-1
```

	Query Match	Similarity	99.8%;	Score 1879;	DB 3;	Length 358;
	Best Local	Similarity	100.0%;	Prod. No. 2,3e-199;		
	Matches	358;	Conservative	0;	Mismatches	0;
					Indels	0;
					Gaps	0;
QY	1	MAPQNLSTFCLLLTLYLIGAVIAGRDFFYKILGVPBSASIKIDIKKAYRKIALQLHPDRNPDD	60			
Db	1	MAPQNLSTFCLLLTLYLIGAVIAGRDFFYKILGVPBSASIKIDIKKAYRKIALQLHPDRNPDD	60			
QY	61	POAOKFODDLCAGAYVYVLSDSERKKQYDYTGEGGLKDGHOSHGDIIFSHPFGPGEFMGCT	120			
Db	61	POAOKFODDLCAGAYVYVLSDSERKKQYDYTGEGGLKDGHOSHGDIIFSHPFGPGEFMGCT	120			
QY	121	PROODRNTIPRGSDIIVDLVEVLTLEEYVAGFVYVNRKPARQAPGRKNCNOEMTTQL	180			
Db	121	PROODRNTIPRGSDIIVDLVEVLTLEEYVAGFVYVNRKPARQAPGRKNCNOEMTTQL	180			
QY	181	GGRGFOQTOEYVVCDECPVNVKLVNERETLEVELEPGVRDQMEPEYLFGEGERPHYDGEFDR	240			
Db	181	GGRGFOQTOEYVVCDECPVNVKLVNERETLEVELEPGVRDQMEPEYLFGEGERPHYDGEFDR	240			
QY	241	FRIKVYKHPFERERRDDLYTNVTSVLVESLVGFEMDITHLDSHKYHISDKITTRPGAXXX	300			
Db	241	FRIKVYKHPFERERRDDLYTNVTSVLVESLVGFEMDITHLDSHKYHISDKITTRPGAXXX	300			
QY	301	KKGECLPFNDNNNIGSLITFDVQPPKQQLNEEAREGIKOLLKQSGYQKVVNNGLOGY	358			
Db	301	KKGECLPFNDNNNIGSLITFDVQPPKQQLNEEAREGIKOLLKQSGYQKVVNNGLOGY	358			

RESULT 4
US-08-868-288A-5
; Sequence 5, Application US/08868288A

```

; CLONE: 306714
us-08-868-288A-5

Query Match      27.4%  Score 515.5;  DB 2;  Length 397;
Best Local Similarity 34.5%  Pred. No. 1.2e+46;
Matches 133;  Conservative 66;  Mismatches 117;  Indels 69;  Gaps 13;

```

RESULT 5
US-09-235-373-5
Sequence 5, Application US/09235373
Patent No. 6001598
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,373
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/868,288
FILING DATE: June 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5: .

ATTORNEY/AGENT INFORMATION:

```

;
MEDIUM TYPE: Floppy disk
COMPILER: IBM PC compatible

```


Query Match	27.3%	Score 515;	DB 2;	length 419;
Best Local Similarity	35.8%;	Pred. No. 1.5e-48;		
Matches 125; Conservative	61;	Mismatches 115;	Indels 48;	Gaps 10

RESULT 8
US-08-974-546-1

Sequence 1, Application US/08974546
Patent No. 5945287
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

Query Match	23.8%	Score 448.5	DB 2	Length 348
Best Local Similarity	30.9%	Pred. No. 2.6e-41		
Matches 119, Conservative	61	Mismatches 100	Indels 105	Gaps 11

RESULT 5

US-08-5/4-540-5
Sequence 5, Application US/08974546
Patent No. 5945287
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
APPLICANT: Shah, Purni
TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.

Page 6

RESULT 10
US-08-472-534-6
: Sequence 6, Application US/08472534
: Patent No. 5919620
: GENERAL INFORMATION:
: APPLICANT: Hamel, Josee

QY	25	DEYKILGPBRSMSIDIKKAYRKALATOLDPRNPDPPOAOKFODLLAAYEVLSDSKRK	84
Db	5	EYDIDLGYSKNNSADEIKKATRKLSKTHPDIIN-KEPABEDKYAEVDBAETLSDDDKRA	63
QY	85	QYDTYGEGBLKDGHOSHSK-----DIFSHFFGDFCFMFGTPEROODRINPR	130
Db	64	AYDYQGAANGANGGEGGAGGFGGFGNAGGFGGFEJLFFSFFGCGSSR--NPNAPR	116
QY	131	-GSDIITVLEVTLEEYVAGNFVEVVRNKPVARQAPGRKRCNCRQEMRTTOLGPRFOMTO	189
Db	117	QGDLDQYRVNLTFFEEALRIGTEKVEKYH--REA-GCRTC-----GSGAKPGTS	161
QY	190	EYVGDCECNVAKLVN-----EXTLE	209
Db	162	PYTCGRCHGAGAVINDTQPLGMMRQVTCVCHGRGKEIKYPCSTTGHGNEKQASHV	221
QY	210	VEIEGVAGDGMAYEPFIGEBEPHVGEPR-GDLRFRIKVVKNHIFERRDDLTYNVTAYVE	268
Db	222	VKIPAGVETGGQINLAAOGGEGAFNGPRGDLVYVSVASBCKFRREGTITFFYNLNLNFVO	281
QY	269	SLVGFEMDITLHDGKHVHISDKITTRPGAXXMKKGEGILPNFDNNINIGSLITFEDVDFPK	328
Db	282	AALGCTVDIPVHG-DVELVPESTQYGGKPKRLSKCAPSLRGAV-GDYLVVNVYTP-	339
QY	329	EQLTEAREGRIKO	341
Db	339	TGLNDROKVALKE	351

RESULT 11
US-08-868-288A-3


```

?      NAME: Billings, Lucy J.
?      REGISTRATION NUMBER: 36,749
?      REFERENCE/DOCKET NUMBER: PF-0309 US
?      TELECOMMUNICATION INFORMATION:
?      TELEPHONE: 415-855-0555
?      TELEFAX: 415-845-4166
?      INFORMATION FOR SEQ ID NO: 7:
?      SEQUENCE CHARACTERISTICS:
?      LENGTH: 277 amino acids
?      TYPE: amino acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      IMMEDIATE SOURCE:
?      LIBRARY: GenBank
?      CLONE: 32470
?      US-08-868-288A-7

Query Match          12.3%; Score 231.5; DB 2; Length 277;
Best Local Similarity 47.6%; Pred. 1.7e-17;
Matches 50; Conservative 14; Mismatches 22; Indels 19; Gaps 3;

QY    26 FYKILGVPRASIKIDIKKKAYRKRIALQLHPDRNPDPQ-AQEKFDOLAAAEVLSDESKR 84
db     4 YETLLOVRASADDIKKIKAYRRKALOWHPDKPKPNKKPFAAKFEKEVAAPAFVYSDTKRPE 63

```

```

Query Match      12.3%; Score 231.5; DB 2; Length 277;
Best Local Similarity 47.6%; Pred. No. 1.7e-17;
Matches 50; Conservative 14; Mismatches 22; Indels 19; Gaps 3

OY 26 FYIILVPSRASIKIDIKKAYRKIALQLHPDRNPDPQ-AQEKFDGLGAAYVLSDSSEKR 84
   :::||| ||||| |||||::||| |||::|||::|||::|||::|||::|||::|||::|||
Db 4 YEILVPSRASADIDIKKAYRKRRALQWHPDKNPDKMKEAEKKFEVAAEVLSDKHRE 63

OY 85 QYVTYGEEL-----KDGHOSSHG-----DIESHPFG 111
   ||| ||| ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 64 YDYGREGSLGTGTGTPSRFAEAGSGGPGFTFTFRSPREVEVREFRFG 108

RESULT 15
US-09-235-373-7
Sequence 7, Application US/09233573
Patent No. 6001598
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRES:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/235,373
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/866,288
FILING DATE: June 3, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-845-4166
TELEFAX: 415-845-0555
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 277 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:11:59 ; Search time 20.55 Seconds
(without alignments)
412.634 Million cell updates/sec

Title: US-09-501-714-1

Perfect score: 1883
Sequence: 1 MAPQNLSTFCLLLYLIGAV.....IKQLKQSGVQKVVNGLCGY 358

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1879	99.8	W94065	Human DnaJ-like pr
2	515	27.3	W00936	Maize DnaJ-related
3	515	27.3	W83397	Maize DnaJ clone p
4	509.5	27.1	Y07061	Renal cancer assoc
5	345	18.3	W22358	S. pneumoniae N-te
6	344.5	18.3	W98475	H. pylori GHPO 885
7	257	13.6	W94066	Human DnaJ-like pr
8	254	13.5	W5710	Homo sapiens fetal
9	242	12.9	W98446	H. pylori GHPO 542
10	239	12.7	Y11105	H. pylori ORF hp7e
11	238.5	12.7	Y11104	H. pylori ORF 03ce
12	216	11.5	W90680	Mouse cysteine str
13	195.5	10.4	W44076	Human secreted pro
14	195.5	10.4	W27640	Secreted protein A
15	194	10.3	W59132	Mus musculus Tubb I
16	188.5	10.0	W36140	Bovine P58 protein
17	151.5	8.0	W87996	A human MCG18 prot
18	146	6.7	Y07029	Breast cancer asso
19	116.5	6.2	W74905	Human secreted pro
20	109.5	5.8	W60599	Sequence of the R1
21	101	5.4	W38514	S. pneumoniae prol
22	100.5	5.3	W31585	Streptococcus pneu
23	100.5	5.3	W88486	Streptococcus pneu
24	99	5.3	W44003	Acidic protease V5
25	94.5	5.0	W94003	H. pylori GHPO 126
26	92	4.9	W98795	H. meningitidis st
27	91	4.8	W93492	N. meningitidis reg
28	90.5	4.8	W19785	Mouse multiple reg
29	89	4.7	W57474	P. falciparum tran
30	88.5	4.7	R77417	Human cell cycle p
31	87.5	4.6	W19784	Human multiple reg
32	86.5	4.6	R05078	Aspergillopepsin A
33	86	4.6	R88058	Protein encoded by
34	86	4.6	R45944	Glutamic acid rece

35	85.5	4.5	953	1	R76707	Recombinant cold-r
36	85.5	4.5	953	1	R90923	F. bidentis pyruva
37	85	4.5	806	1	R27481	RP-III residual pr
38	85	4.5	1482	1	R44193	Rat NMDA receptor
39	83	4.4	589	1	P60303	Sequence encoded b
40	83	4.4	764	1	W26727	Yeast Mps1p protei
41	82	4.4	393	1	W61252	Salicylic acid ind
42	82	4.4	1195	1	R05530	High density lipop
43	82	4.4	1292	1	R05531	High density lipop
44	82	4.4	1456	1	R49042	NMDA receptor chan
45	81.5	4.3	394	1	R35221	Protease B. Procta

ALIGNMENTS

RESULT	ID	Score	Query Match Length	ID	Description
W94065	W94065	99.8	358	1	Human DnaJ-like pr
AC	W94065	99.8	358	1	Human DnaJ-like pr
DT	09-APR-1999	99.8	358	1	Human DnaJ-like pr
DE	Human DnaJ-like protein, HSPJ1.	99.8	358	1	Human DnaJ-like pr
KW	DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;	99.8	358	1	Human DnaJ-like pr
KW	Immune disorder; inflammation; tissue damage; diabetes; wound healing;	99.8	358	1	Human DnaJ-like pr
KW	Chromosome mapping.	99.8	358	1	Human DnaJ-like pr
OS	Homo sapiens.	99.8	358	1	Human DnaJ-like pr
FH	Key	99.8	358	1	Human DnaJ-like pr
FT	Misc-difference 298	99.8	358	1	Human DnaJ-like pr
FT	/label= unknown	99.8	358	1	Human DnaJ-like pr
FT	/note= "encoded by AAN"	99.8	358	1	Human DnaJ-like pr
FT	Misc-difference 299	99.8	358	1	Human DnaJ-like pr
FT	/label= unknown	99.8	358	1	Human DnaJ-like pr
FT	/note= "encoded by TAN"	99.8	358	1	Human DnaJ-like pr
PN	W0985509-A2.	99.8	358	1	Human DnaJ-like pr
PD	10-DEC-1998.	99.8	358	1	Human DnaJ-like pr
PE	02-JUN-1998; U11182.	99.8	358	1	Human DnaJ-like pr
PR	03-JUN-1997; US-868288.	99.8	358	1	Human DnaJ-like pr
PA	(TNCY-) INCYTE PHARM INC.	99.8	358	1	Human DnaJ-like pr
PI	Au-Young J, Bandman O, Lal P;	99.8	358	1	Human DnaJ-like pr
DR	WPI; 99-070259/06.	99.8	358	1	Human DnaJ-like pr
DR	N-PSDB; X06099.	99.8	358	1	Human DnaJ-like pr
PT	New nucleic acid encoding human DnaJ-like proteins - for diagnosis,	99.8	358	1	Human DnaJ-like pr
PT	treatment and prevention of cancer, immune disorders and	99.8	358	1	Human DnaJ-like pr
PT	Inflammation	99.8	358	1	Human DnaJ-like pr
PS	Claim 1; Fig 1A-D; 73pp; English.	99.8	358	1	Human DnaJ-like pr
CC	This represents a human DnaJ-like protein, HSPJ1. The invention provides	99.8	358	1	Human DnaJ-like pr
CC	two human DnaJ-like proteins which are heat shock proteins J1 and J2	99.8	358	1	Human DnaJ-like pr
CC	(HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host	99.8	358	1	Human DnaJ-like pr
CC	cells containing a vector comprising the nucleic acids are used for the	99.8	358	1	Human DnaJ-like pr
CC	production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used	99.8	358	1	Human DnaJ-like pr
CC	to raise Ab, therapeutically and to screen for specific binding agents.	99.8	358	1	Human DnaJ-like pr
CC	Antagonists are used to treat or prevent a wide variety of solid cancers,	99.8	358	1	Human DnaJ-like pr
CC	leukemia and lymphoma; immune disorders (typical of many disclosed are	99.8	358	1	Human DnaJ-like pr
CC	acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,	99.8	358	1	Human DnaJ-like pr
CC	diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial	99.8	358	1	Human DnaJ-like pr
CC	and parasitic infections) and inflammation. Agonists may be used to treat	99.8	358	1	Human DnaJ-like pr
CC	or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart	99.8	358	1	Human DnaJ-like pr
CC	attack, ischemia, UV exposure, insulin-dependent diabetes and for wound	99.8	358	1	Human DnaJ-like pr
CC	healing, and may be expressed from a vector containing the nucleic acids.	99.8	358	1	Human DnaJ-like pr
CC	Fragments of the nucleic acids are used as primers and probes for	99.8	358	1	Human DnaJ-like pr
CC	detecting and quantifying the HSPJ-encoding nucleic acid in usual	99.8	358	1	Human DnaJ-like pr
CC	hybridisation and/or amplification assays, therapeutically as antisense,	99.8	358	1	Human DnaJ-like pr
CC	triplex-forming or ribozyme molecules, and for chromosome mapping.	99.8	358	1	Human DnaJ-like pr
SQ	Sequence 358 AA;	99.8	358	1	Human DnaJ-like pr
Query Match	99.88; Score 1879; DB 1; Length 358;	99.88	358	1	Human DnaJ-like pr
Best Local Similarity	100.0%; Pred. No. 6,7e-184;	100.0	358	1	Human DnaJ-like pr
Matches 358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		358	358	1	Human DnaJ-like pr
OY	1 MAPQNLSTFCLLLYLIGAVIAGRDFYKIGVPSASIKDKKRYKRLAQLHPDRPDD 60				
DB	1 MAPQNLSTFCLLLYLIGAVIAGRDFYKIGVPSASIKDKKRYKRLAQLHPDRPDD 60				

```

QY 61 PQAQEKFDLGAAYEVLSDSEKRAYDYTYGEGGLKDGSHSGDIFSHFDFGFMFGCT 120
DB 61 PQAQEKFDLGAAYEVLSDSEKRAYDYTYGEGGLKDGSHSGDIFSHFDFGFMFGCT 120
QY 121 PRODRNIPRSDIIVDLEVTLEEVYAGNFEVYVNRKPVARQAPGRKRCNCRQEMRTTOL 180
DB 121 PRODRNIPRSDIIVDLEVTLEEVYAGNFEVYVNRKPVARQAPGRKRCNCRQEMRTTOL 180
QY 181 GPGFQMTQEVVCECPNVKLVNEERTLEVEIEBVGVDGMETPTIGGEPHVDDEPDRL 240
DB 181 GPGFQMTQEVVCECPNVKLVNEERTLEVEIEBVGVDGMETPTIGGEPHVDDEPDRL 240
QY 241 FRIVVKNHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGHRKHSRDKITPRGAXXW 300
DB 241 FRIVVKNHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGHRKHSRDKITPRGAXXW 300
QY 301 KKGGLNPNFNNIKGSLITTFDVPFKEQLTEPARGICQLKQSGVQKVVNGLQGY 358
DB 301 KKGGLNPNFNNIKGSLITTFDVPFKEQLTEPARGICQLKQSGVQKVVNGLQGY 358

RESULT 2
W00936
ID W00936 standard; Protein; 419 AA.
AC W00936;
DT 11-NOV-1997 (first entry)
DE Maize DnaJ-related protein.
KW DnaJ-related protein; chaperone; protein folding; DNA replication;
KW translation; peptide translocation; ZMDJ1; promoter; maize;
KW transgenic plant; insecticide; antifungal; fungicide;
KW crop protection.
OS Zea mays L. cv. B73.
PN W09705260-A2.
PD 13-FEB-1997.
PF 12-JUL-1996; U11676.
PR 26-JUL-1995; US-001522.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczyński C, Horowitz J, Rosichan JL;
DR N-PSDB; T84330.
DR WPI; 97-145697/13.
PT Tobacco DnaJ-related gene transcription/translation regulatory
PT sequence, ZMDJ1 - is intermediate between constitutive and tissue
PT specific promoters, partic. for control of antifungal and
PT insecticide genes
PS Example 1; Fig 2A-B; 26pp; English.
CC This polypeptide sequence is encoded by the maize DnaJ-related
CC gene (see T84330). DnaJ-related proteins assist in chaperone-
CC mediated protein folding and provide cell viability at high
CC temperatures. They are also involved in DNA replication,
CC translation and peptide translocation across intracellular
CC membranes. Due to this wide range of functions, DnaJ has a wide
CC range of effectiveness and the gene's promoter sequence, ZMDJ1
CC (see T84329), is effective in a wide range of tissues. The
CC claimed ZMDJ1 promoter can be used in claimed methods for control
CC of antifungal or insecticidal genes in transgenic plants.
SQ Sequence 419 AA.

Query Match 27.3%; Score 515; DB 1; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.1e-44;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

```

```

DB 127 VSLIEDLYNGTSSKRLSLRNVCCKGKSGASMRCPGCGSGAKVTIR--QLGPSMT 184
QY 186 QMTQEVVCECPNV-----KLVNEERTLEVEIEPGVROMEXPTIGEGF 229
DB 185 QOMQO-PCNCKGTGESINEKRCPCGKGKGVIOEKVLEVHVEKGMQHNQKITPFGEND 243
QY 230 PHVDGERGDLRFPRIKVKNHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLDGKHYHSR 289
DB 244 EAPDYTGVDLYFVYQOKDSKFRKGGEDLYERTLSTLTXLCSFQVTLHDMRQLLINS 303
QY 290 D--KITRPGAXXMKKGLPNFNNIKGSLITTFDVPFKEQLTEPAR 336
DB 304 DPGEVVQDQFKALINDEGMPIYQRPFMKGLYHFTVEEPDLSLAPQCK 352

```

```

RESULT 3
W83397
ID W83397 standard; Protein; 419 AA.
AC W83397;
DT 23-FEB-1999 (first entry)
DE Maize DnaJ clone protein sequence.
KW Maize; Zea mays; transcriptional promoter; control sequence; ZMDJ1;
KW DnaJ clone; corn; insecticide; antifungal protein; transgenic plant;
KW insect resistance; European corn borer; fungal resistance;
KW Ostrinia nubilalis.
OS Zea mays.
PN US5850018-A.
PD 15-DEC-1998.
PF 26-JUL-1996; 686417.
PR 27-JUL-1995; US-001552.
PR 26-JUL-1996; US-686417.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczyński C, Horowitz J, Rosichan JL;
DR WPI; 99-069853/06.
DR N-PSDB; V72905.
PT Maize ZMDJ1 promoter - and corresponding recombinant expression
PT systems and transgenic plants
PS Example 1; Fig 2; 16pp; English.
CC The present invention describes the maize ZMDJ1 control sequence.
CC described are: (1) a recombinant expression system comprising the maize
CC ZMDJ1 control sequence linked to a heterologous protein coding sequence;
CC (2) a monocotyledonous plant, plant part or plant cell modified to
CC contain the expression system; (3) DNA comprising fragments of the ZMDJ1
CC control sequence which retain transcription-initiating activity and/or
CC the function of the leader sequence; and (4) a method to regulate the
CC expression of a gene under control of the ZMDJ1 control sequence in
CC plant cell, plant parts or plants, where the plant is a monocotyledon,
CC comprising modifying the cell, part or plant containing the gene under
CC control of the ZMDJ1 control sequence to contain the DNA of (3) or RNA
CC of the same nucleotide sequence. Expression systems as described above
CC in which the heterologous sequence encodes an insecticidal or antifungal
CC protein can be used to produce transgenic maize plants resistant to
CC insects (e.g. European corn borer) or fungi (e.g. Ostrinia nubilalis).
CC The control sequence comprises a constitutive promoter that directs
CC transgene expression in various plant tissues, including 11-week-old
CC leaf blades, leaf whorls, leaf collars, stalk rind, stalk pith, stalk
CC nodes, roots and kernels. The present sequence represents the protein
CC encoded by a maize DnaJ clone from the present invention.
SQ Sequence 419 AA.

```

```

Query Match 27.3%; Score 515; DB 1; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.1e-44;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

```


QY 140 VTLSEVYAGNFVEVVRNKPV-----AROAPGRKCNCRQEMRTTOLGPREF 185
 Db 127 VSLFDLITNGSKLISLRNWKCSKCKGSKGSMRCPCGQSGMKVTLR--OLGPMI 184
 QY 186 QMTQEVVCDRCPNV-----KLVEERTLEVEIEIPGVADGMEYPIGEGE 229
 Db 185 QQMQQ--PCNCKGKGSINKEKDRCPGCKGEKIVLEKVLVEYHVKGMQHNCKITFPGRAD 243
 QY 230 PHVDGEGDLRFRTKYVKKHPIFERKDDLYTNVTVSLVESLVGFEMDITLHDKVHISR 289
 Db 244 EAPDTVGDIVFVLOQDHSKFKRKGEDLFYEHTLSLLEALCGFQVLTLDNRQLIKS 303
 QY 290 D--KITRPGAXXMKKGGELPNEDNNIKSLITFEDVEPEQLTEEAR 336
 Db 304 DPGEVVAKPDQFKALNDEGMPIYQRPFKMKLYIHTEVFPSLAEPOCK 352

RESULT 4

ID Y07061 standard; Protein; 415 AA.
 AC Y07061;
 DT 02-JUL-1999 (first entry)
 DE Renal cancer associated antigen precursor sequence.
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KM breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KM prostate cancer.
 OS Homo sapiens.
 PN W0904265-A2.
 PD 28-JAN-1999.
 PF 15-JUL-1998; U14679.
 PR 22-JUN-1998; US-102322.
 PR 17-JUL-1997; US-896164.
 PR 10-OCT-1997; US-061599.
 PR 10-OCT-1997; US-061765.
 PR 10-OCT-1997; US-948705.
 PR 11-OCT-1997; GB-021697.
 PA (LUDWIG) LUDWIG INST CANCER RES.
 PI Chen Y, Gout I, Gure A, Ohare M, Obata Y, Old LJ,
 PI Freundschuh M, Sahin U, Scanlan MJ, Stockert E,
 PI Tureci O;
 DR WPI; 99-132448/11.
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers
 PS Disclosure; Page 456-457; 787pp; English.
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 SQ Sequence 415 AA;

Query Match 27.1%; Score 509.5; DB 1; Length 415;
 Best Local Similarity 37.0%; Pred. No. 7.6e-44;
 Matches 128; Conservative 55; Mismatches 122; Indels 41; Gaps 12;

QY 19 AVINGRDFYKILGVPKRSASIKDIKAYRKLALQHPDRNPDDQAQE-KPQDLCAAYVYL 77
 Db 2 ANVADTKLYLDLGVPAAGSENEMLKAYRKLAKKEYHPDKN---PQMQETNFEISFAEVYL 58
 QY 78 SDSEKRRQDYTVYGEGLKDGHOSS--HGDIPI--HFGDGFPMFGGTTPROODRNIPRGS 132
 Db 59 SNPKRELIDYKYGQGLREGSSGGGWHGLIFSLVFCGLFGFM--GNQSSRNORRRGE 116

QY 133 DIIVDEVTLEEVYAGNFVEVVRNKPVARQA-----PKR-----RKC-N-CR---QEMRTTOL 180
 Db 117 DMHPLKVLSLDLNKGKTKTLQLSKNVLCASCSGQGGSGAVQKCSACRGRVIMLRQL 176
 QY 181 GPGRFQMTQEVV-----CDECPNVKLVEERTLEVEIEIPGVADGMEYPI 225
 Db 177 APGAVQMQSVCSDCNGGEVYINKEKDRCKCEGKVKIYEVILEYHVDKMKHQRIFTT 236
 QY 226 GEGEPHVDGEGDLRFRTKYVKKHPIFERKDDLYTNVTVSLVESLVGFEMDITLHDKH 284
 Db 237 GEADQAPEMNBTLEFLLPGEKMEVFQRDGNDLHMVTKILVLAICGFQFTLSHLDGRQ 296
 QY 285 VHTS--RDKITRPGAXXMKKGGELPNEDNNIKSLITFVDPVPK 328
 Db 297 IVVXYPGKVIIEPGCVRVVRGEMPOYRNPEKGLYIKFVDPPE 342

RESULT 5

ID W22358 standard; Protein; 352 AA.
 AC W22358;
 DT 01-OCT-1997 (first entry)
 DE S. pneumoniae N-terminal portion of DnaJ protein.
 KW Streptococcus pneumoniae; S. pyogenes; S. agalactiae; genome; HSP72;
 KM open reading frame; heat shock protein 72; DnaJ; DnaK; chimeric;
 KW E. coli; Lactococcus lactis; antibody; vaccine; infection; human.
 OS Streptococcus pneumoniae.
 PN W09640928-A1.
 PD 19-DEC-1996.
 PF 17-MAY-1996; CA0322.
 PR 07-JUN-1995; US-472534.
 PR 04-AUG-1995; US-001805.
 PA (IAFB-) IAF BIOVAC INC.
 PI Brodeur B, Hamel J, Martin D, Rioux C;
 DR WPI; 97-052328/05.
 DR N-PSDS; T73388.
 PT Streptococcal heat shock proteins and corresponding DNA sequences -
 PT used in the production of a vaccine to treating and preventing
 PT strain-specific Streptococcal infection
 PS Example 3; Page 93-94; 156pp; English.
 CC This amino acid sequence corresponds to the N-terminal portion of the
 CC Streptococcus pneumoniae DnaJ protein. The protein shows a high degree
 CC of identity (72% and 51%) to the DnaJ proteins from Lactococcus lactis
 CC and E. coli respectively. The sequence encoding this truncated protein
 CC was isolated on the same nucleotide sequence that encodes the
 CC S. pneumoniae heat shock protein 72 (HSP72; W22357). The nucleotide
 CC fragment was isolated from a HindIII-partially digested genomic DNA
 CC library using a fragment of the chimeric gene (T73392) corresponding to
 CC the sequence encoding the C-terminal 169 amino acids of HSP72. The HSP72
 CC protein and its fragment, or antibodies specific to HSP72, are used in
 CC pharmaceutical compositions, pref. a vaccine, for treating or preventing
 CC infection by S. pneumoniae or related bacteria in humans, e.g.
 CC S. pyogenes or S. agalactiae.
 SQ Sequence 352 AA;

Query Match 18.3%; Score 345; DB 1; Length 352;
 Best Local Similarity 28.4%; Pred. No. 3.9e-27;
 Matches 106; Conservative 55; Mismatches 130; Indels 82; Gaps 13;

QY 25 DFKILGVPRASIKDIKAYRKLALQHPDRNPDDQAQEFQDLCAAYEVLSDSEKRR 84
 Db 5 EBYRLGVSKNNADEIKAYRKLKSKYHPDIN-KEPAEKKYKVEVQAYETLSDQKRA 63
 QY 85 QYDTYGEGLKDGHOSSHG-----DIFSHFPDGFPMFGGTTPROODRNIPR 130
 Db 64 AYDQYGAANGAGGAGGAGGFGGAGGFGGDFETFSFVFG-----GSSR--NPNAIPR 116
 QY 131 -GSDIIVDEVTLEEVYAGNFVEVVRNKPVARQAAGKRCCKRQEMRTTOLGPGFQMTQ 189
 Db 117 QGDDLDQYVNVNLFPEALIGTEKEVYH-----REA-GCRTCN-----GSAKAPETS 161

```

OY 190 EVVDECPNVLVN-----EERTLE 209
DB 162 FVTGCRGAGVINVDTQPLGMRRQVTCVCHGRKEIKYPCCTGTCGHEQASHVH 221
OY 210 VEIRPGVADGNEYPIFGGEPHNGEP-GDLRFRIKYVKHPIFERRGDLYTNVTSLVE 268
DB 222 VKIAGVETGQOIRLAGGEGAGFNGPGDLYVYVSVEASDKFEREGTTFIYNLNLNFVQ 281
OY 269 SLVGFEMDITLHDGKHVHISRDKITRPGAXXWKKGEGLPNNNNINGSLITFDVDEPK 328
DB 282 AALGDYDIPYVHG-DVELVPECTGTGKRRKRLSKAPSLRGAV-GDOYTVNVVTP- 339
OY 329 EQLTEAREGKIQ 341
DB 339 TGLNDRQKVALKE 351

```

```

RESULT 6
W98475 ID W98475 standard; Protein; 369 AA.
AC W98475;
DT 31-MAR-1999 (first entry)
DE H. pylori GHPO 885 protein.
KM GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
KM peptic ulcer disease.
ON Helicobacter pylori.
PN W09843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PI A1-Garawi A, Kleantous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.
DR N-PSDB: X14194.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 853-855; 2054PP; English.
CC This sequence represents a Helicobacter pylori GHPO protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 369 AA;

```

```

Query Match 18.3%; Score 344.5; DB 1; Length 369;
Best Local Similarity 28.7%; Pred. No. 4,6e-27;
Matches 100; Conservative 64; Mismatches 143; Indels 41; Gaps 10;
OY 26 FYKILGVPKRSKDKIKKRYKRLALQLHPDRNDPQADKQDGLAAEVLVSDSKRKQ 85
DB 5 YVEILFEVHSHQETKTKRYLALHPDRNAGDEAEKEKRLKLINEAVLSDKKRAL 64
OY 86 YDTYGEGLKDKHSHSGDIFSHFGD-----FGFAGGTGPRQODNIRGSDIYD 137
DB 65 YDRYCKKLNQA-GASQGD-FSDPFEDLGSFEDAFGFAARSKKQSSIAIP---DYLOT 119
OY 138 LEVTLAEVYAGNFVEY-VANKPVANQAPK-----RKCNCROEMRTTQ----- 180
DB 120 LELSKENAFVGCCKTKITKYQGVCSGCDGTAKDKALETCQNGCNGQGVFMQGFMSFAQ 179
OY 180 -----LGRGFQMTQEVWDECPNKLIVNEERTLEVEIRPGVADGMEYPIFGGEPHYDG 234
DB 180 TCGAOCGKGYKTP---CQACKGTYILKDEIDALIPGIDDNRMVLRKKNKNEYEKG 236
OY 235 EFGDLRFRIKYVKHPIFERRGDLYTNVTSLVEVGFEMDITLHDGKHVHISDKITR 294
DB 235 EFGDLRFRIKYVKHPIFERRGDLYTNVTSLVEVGFEMDITLHDGKHVHISDKITR 294

```

```

DB 237 KRGLDLYEAQYKEDHEHRRKRCGLFLFAVPVFTTIALGHTIKVPSLKGDELLEKIPRNAR 296
OY 295 PGAXXWKKGGLPFPNDNNNIKSLITFDVDFPREQJTEAREGRIKOL 342
DB 297 DKQTFARNEGKVKH-PSSYRGLSLVELQVITYRK-SLNRKQGLLEKRL 342
RESULT 7
W94066 ID W94066 standard; Protein; 330 AA.
AC W94066;
DT 09-APR-1999 (first entry)
DE Human DnaJ-like protein, HSPJ2.
KM DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
KM immune disorder; inflammation; tissue damage; diabetes; wound healing;
OS Homo sapiens.
FH Key Location/Qualifiers
FH Misc_difference 239
FT /label= unknown
FT /note= "encoded by GST"
FT Misc_difference 272
FT /label= unknown
FT /note= "encoded by GNG"
FT Misc_difference 290
FT /label= unknown
FT /note= "encoded by GNG"
FT Misc_difference 291
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc_difference 293
FT /label= unknown
FT /note= "encoded by NNC"
FT Misc_difference 298
FT /label= unknown
FT /note= "encoded by NCA"
FT Misc_difference 301
FT /label= unknown
FT /note= "encoded by TTN"
FT Misc_difference 302
FT /label= unknown
FT /note= "encoded by NNA"
FT Misc_difference 305
FT /label= unknown
FT /note= "encoded by NCC"
FT Misc_difference 309
FT /label= unknown
FT /note= "encoded by NNA"
W09855509-A2.
PD 10-DEC-1998.
PF 02-JUN-1998; U11182.
PR 03-JUN-1997; US-868288.
PA (INCY-) INCYTE PHARM INC.
PI Au-Young J, Bandman O, Lal P;
DR WPI: 99-070259/06.
DR N-PSDB: X06100.
PT New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
PT treatment and prevention of cancer, immune disorders and
PT inflammation
PS Claim 22; Fig 3A-D; 73PP; English.
CC This represents a human DnaJ-like protein, HSPJ2. The invention provides
CC two human DnaJ-like proteins which are heat shock proteins J1 and J2
CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
CC cells containing a vector comprising the nucleic acids are used for the
CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
CC to raise Ab, therapeutically and to screen for specific binding agents.
CC Antagonists are used to treat or prevent a wide variety of solid cancers,
CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC and parasitic infections) and inflammation. Agonists may be used to treat
CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound

```

Y11105
ID Y11105 standard; Protein; 278 AA
AC Y11105;

Query Match	12.7%	Score 239	DB 1	Length 278
Best Local Similarity	25.4%	Pred. No. 1.9e-16		
Matches 82	Conservative	117	Indels 72	Gaps 10

OY	24	RDEYIIIGVPSASIKDKIKKARYKATLQIHIDRNDDPQAOKRFODLGAAEVLSDSEKR	83
Dd	3	KSLVTTLAVSENASODELKSTRIRAKRYHPLN-KTKEAEKKFEINAAVEILDSDEKR	61
OY	84	KQYDTYGE-----EGLKD-----GHOSHGDIFFSHFPGDFGFMFGTPRQQ-----DR	126
Dd	62	KQYDQFGDNMGEGGNFSDFARSRGSEDLDDLSSIFPKGGESQGFNSQGSFGSFPNSN	121
OY	127	NIRPGSDIIVLEVTLYEEVYAGNEVEYVRKNRPVARQAPGRKRCNCROEMRTTQLGPGRFQ	186
Dd	122	FAPENLDATALINVSVDLTLLGNKKQV-----	149
OY	187	MTEGVCEDECNNVKLVINEERTLEVIEEPGYVDGMETPIEGEGPHVDGPDGLRRRIKYV	246
Dd	149	-----SVNNE-TFSLKITPIGVGEGERIVRRNKGMKGTRG-GDLLQIHIE	192
OY	247	KHPFERRGDLYNVTVSLVESLVGFMEDITHLDGHK-VHISRDKITRPGAXXWKKEGG	305
Dd	193	EDEMYYRKREKDIIQIFDLPLKTALFSGKIETA--TWHTLTLLTIPPNKKAQMKFKIKDGG	250
OY	306	LPNEDNNNIKGSLLITEPVDEPPK	328
Dd	251	IKSRKTSHVGDCLASSFDL-PK	271
 RESULT 11 Y11104			
ID	Y11104	standard; Protein; 275 AA.	
AC	Y11104;		
DC	08-JUN-1999	(first entry)	
DE	H. pylori ORF 05ce10910.23712780.c1.4	cytoplasmic protein.	
DE	Vaccine: probe: diagnostic: ORF: cell envelope protein;		
KW	secreted protein; cytoplasmic protein; cellular protein.		
OS	Helicobacter pylori.		
PN	WO9824475-A1.		
RD	11-JUN-1998.		

Query Match	12.78;	Score 238.5;	DB 1;	Length 275;
Best Local Similarity	26.98;	Pred. No. 2.1e-16;		
Matches 73; Conservative	39;	Mismatches 92;	Indels 67;	Gaps 7;

QY	24	RDEYKLTIGPVRASLTKDITKKAYRKLATQLPDPNPDPPOAKSFODDIAAYEYLSDSEKR	83
Db	3	KSLYQVLANSNAASODELTKRYRLAQIYPDLN-KTKAEKFEKINNAEYLLSDSEKR	61
QY	84	KQYDTYGE-----EELKD-----GHOSHGDIFSHFFGDFGFMFGGTPRQO-----DR	126
Db	62	RQDQFGDMFGGQNFSPFASRSPSEDLDTLISIFGKGFSQSPFSONSGFSGFNFPSN	121
QY	127	NIPRGSDIIVDLEVLVEVYAGNVEVYVRNKKPVARQAPKRCNCNRQMRPTQLGPPRFQ	166
Db	122	FAPENLDYALINLSVLDITLLGNKKQV-----	149
QY	187	MNOEYVCCDECPWKLVDNEKRLVELEPDPGVDGMEYPIIGSECPHYDGPDLFRIRIKV	246
Db	149	-----SVNNE-TFSLKLPVIGEEBCKIRVRNKKGMKRTGR-GDLLLIQIHIE	192
QY	247	KHPIFERRGDDLYTNVTVSLVESLVEGMDI	277
Db	193	EDEMVRREKDDITQIFDLPKTLAFGGKIEI	223
RESULT	12		
R90680	ID	R90680 standard; Protein; 223 AA.	
AC	R90680;		
DT	07-APR-1996	(first entry)	
DE	Mouse	cysteine string protein.	
KW	Mouse;	cysteine string protein; anergy; T-lymphocyte;	
KW	different	display; cDNA; reverse transcription; probe;	
KW	polymerase	chain reaction; cloning; gel electrophoresis;	
KW	antibody;	diagnostic; immunassay; autoimmune disease;	
KW	organ	transplantation; vaccine; drug screening.	
OS	Mus	musculus.	
FT	Key	Location/Qualifiers	
FT	misc_difference	168	
FT	/note	"In-frame stop codon"	
FT	misc_difference	183	
FT	/note	"In-frame stop codon"	
FT	misc_difference	191	
FT	/note	"In-frame stop codon"	
PD	WO9600300-A1.		
PD	04-JAN-1996.		
PD	22-JUN-1995;	U07958.	

FT	Misc_difference	110	/label=	Unspecified
FT			/note=	"encoded by Gyr"
FT	Misc_difference	111	/label=	Unspecified
FT			/note=	"encoded by NKC"
FT	Misc_difference	112	/label=	Unspecified
FT			/note=	"encoded by NKr"
FT	Misc_difference	113	/label=	Unspecified
FT			/note=	"encoded by TMN"
FT	Misc_difference	114	/label=	Unspecified
FT			/note=	"encoded by YTr"
FT	Misc_difference	115	/label=	Unspecified
FT			/note=	"encoded by YMA"
FT	Misc_difference	116	/label=	Unspecified
FT			/note=	"encoded by YTA"
FT	Misc_difference	117	/label=	Unspecified
FT			/note=	"encoded by YTr"
FT	Misc_difference	118	/label=	Unspecified
FT			/note=	"encoded by TSA"
FT	Misc_difference	119	/label=	Unspecified
FT			/note=	"encoded by TGA a stop codon"

	/note= "encoded by TAA a stop codon "
Misc_difference	128
FT	/label= Unspecified
FT	
FT	

ET	Misc_difference	129	/label=	Unspecified
ET			/note=	encoded by ARR
ET	Misc_difference	130	/label=	Unspecified
ET			/note=	encoded by CVA
ET	Misc_difference	134	/label=	Unspecified
ET			/note=	encoded by CVA

FT	Misc_difference	135	/label=	Unspecified	encoded by NNA
FT			/note=	encoded by YAA	
FT	Misc_difference	136	/label=	Unspecified	encoded by NNA
FT			/note=	encoded by YAA	
FT	Misc_difference	137	/label=	Unspecified	encoded by YKT
FT			/note=	encoded by YKT	
FT	Misc_difference	139	/label=	Unspecified	encoded by GRR
FT			/note=	encoded by GRR	
FT	Misc_difference	140	/label=	Unspecified	encoded by RWR
FT			/note=	encoded by RWR	
FT	Misc_difference	141	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	142	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	143	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	144	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	145	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	146	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	147	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	148	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	
FT	Misc_difference	149	/label=	Unspecified	encoded by WYC
FT			/note=	encoded by WYC	

FT /label= Unspecified
FT /note= "encoded by SYC"
FT Misc_difference 151
FT /label= Unspecified
FT /note= "encoded by GNG"
FT Misc_difference 155
FT /label= Unspecified
FT /note= "encoded by TGN"
PN W09739123-A2.
PD 23-OCT-1997;
PR 14-APR-1997; U06139.
PR 18-APR-1996; US-634325.
PA (GEMV) GENETICS INST INC.
PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racie LA,
PI Spaulding V;
DR WPI: 97-526460/48.
DR N-PSDB: V02136.
PT New secreted proteins encoded clones present in ATCC 98026 -
PT possibly having cytokine, cell proliferation/differentiation
PT regulating, immunomodulating and many other activities
PS Claim 12; Page 78-79; 139pp; English.
CC The present sequence represents a novel human secreted protein deposited
CC under accession number ATCC 98026. The secreted protein can be used to
CC determine biological activity, to raise antibodies, as tissue markers,
CC to isolate cognate ligands or receptors, to identify agents that
CC modulate their interactions and as nutritional supplements. It may also
CC have a very wide range of biological activities although no evidence
CC for any is provided in the specification. Typical of these are cytokine,
CC cell proliferation/differentiation modulating activity or induction of
CC other cytokines; immunostimulating/immunosuppressant activities (e.g.
CC for treating human immunodeficiency virus infection, cancer, autoimmune
CC diseases and allergy); regulation of haematopoiesis (e.g. for treating
CC anaemia or as adjunct to chemotherapy); stimulation of growth of bone,
CC cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds,
CC periodontal disease, neurological diseases stroke, fibrosis); inhibition
CC or stimulation of follicle stimulating hormone (for control of
CC fertility); chemotactic and chemokinetic activities (e.g. for treating
CC infections, tumours); haemostatic or thrombolytic activity (e.g. for
CC treating haemophilia, cardiac infarction etc.); anti-inflammatory
CC activity (e.g. for treating septic shock, Crohn's disease); as
CC antimicrobials; for treating psoriasis or other hyperproliferative
CC disease; for regulation of metabolism, behaviour, and many others. Also
CC contemplated is the use of the corresponding nucleic acid in gene
CC therapy procedures.
SQ Sequence 159 AA;

Query Match 10.4%; Score 195.5; DB 1; Length 159;
Best local Similarity 43.4%; Pred. No. 2.3e-12;
Matches 36; Conservative 21; Mismatches 25; Indels 1; Gaps 1;

QY 3 PQLNSTFCLLLVIGAVINGRDFYKILGVPKRSASIKDIKAYRKLALQHPDPDPQ 62
DB 4 PGIETFAICILMTELILASKSYDILGVPKRSASERQIKAKHKLAKMKHPDKN-KSPD 62

QY 63 AQKFPDDLGAAYVLSDEKRRQ 85
DB 63 AEAKFREIAEAYETLSDANXTKR 85

RESULT 14
W27640
ID W27640 standard; Protein: 159 AA.
AC W27640;
DT 14-MAY-1998 (first entry)
DE Secreted protein AM610.
KW Human; secreted protein; research; treatment; AM610.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Misc_difference 82 /label= unknown
FT /note= "encoded by AGN"
FT Misc_difference 87

FT /label= unknown
FT /note= "encoded by TGA"
FT Misc_difference 96
FT /label= unknown
FT /note= "encoded by TAG"
FT Misc_difference 102
FT /label= unknown
FT /note= "encoded by RGR"
FT Misc_difference 103
FT /label= unknown
FT /note= "encoded by RGT"
FT Misc_difference 106
FT /label= unknown
FT /note= "encoded by RAG"
FT Misc_difference 109
FT /label= unknown
FT /note= "encoded by TGA"
FT Misc_difference 110
FT /label= unknown
FT /note= "encoded by GYR"
FT Misc_difference 111
FT /label= unknown
FT /note= "encoded by NKC"
FT Misc_difference 112
FT /label= unknown
FT /note= "encoded by NKT"
FT Misc_difference 113
FT /label= unknown
FT /note= "encoded by TMN"
FT Misc_difference 114
FT /label= unknown
FT /note= "encoded by YTR"
FT Misc_difference 115
FT /label= unknown
FT /note= "encoded by YAA"
FT Misc_difference 116
FT /label= unknown
FT /note= "encoded by YTA"
FT Misc_difference 117
FT /label= unknown
FT /note= "encoded by TSA"
FT Misc_difference 118
FT /label= unknown
FT /note= "encoded by TGA"
FT Misc_difference 121
FT /label= unknown
FT /note= "encoded by TAA"
FT Misc_difference 128
FT /label= unknown
FT /note= "encoded by TYN"
FT Misc_difference 129
FT /label= unknown
FT /note= "encoded by ARR"
FT Misc_difference 130
FT /label= unknown
FT /note= "encoded by CYA"
FT Misc_difference 132
FT /label= unknown
FT /note= "encoded by CAY"
FT Misc_difference 134
FT /label= unknown
FT /note= "encoded by AKC"
FT Misc_difference 135
FT /label= unknown
FT /note= "encoded by YAA"
FT Misc_difference 136
FT /label= unknown
FT /note= "encoded by NAA"
FT Misc_difference 137
FT /label= unknown
FT /note= "encoded by YKT"
FT Misc_difference 139
FT /label= unknown
FT /note= "encoded by GRR"

```
FT Misc_difference 140 /label= unknown
FT /note= "encoded by RWI"
FT Misc_difference 141 /label= unknown
FT /note= "encoded by CAW"
FT Misc_difference 142 /label= unknown
FT /note= "encoded by WYC"
FT Misc_difference 143 /label= unknown
FT /note= "encoded by CNN"
FT Misc_difference 146 /label= unknown
FT /note= "encoded by NNN"
FT Misc_difference 147 /label= unknown
FT /note= "encoded by WKG"
FT Misc_difference 148 /note= "encoded by GTK"
FT Misc_difference 149 /label= unknown
FT /note= "encoded by SYC"
FT Misc_difference 151 /label= unknown
FT /note= "encoded by GNG"
FT Misc_difference 155 /label= unknown
FT /note= "encoded by TGN"
FT W09739122-AZ.
PD 23-OCT-1997.
PF 11-APR-1997; U06042.
PR 12-APR-1996; US-631184.
PA (MUR0-) MUR0 PHARM INC.
PI Theoharides TC;
DR WPI: 97-526459/48.
DR N-PSDB: T88058.
PT Human and murine secreted proteins - useful to research or treat
PS diseases or disorders related to their function
PS Claim 11; Pages 78-79; 140pp; English.
CC The present sequence is a human secreted protein, which may
CC have nutritional uses, or cytokine and cell
CC proliferation/differentiation, immune stimulating or suppressing,
CC haematopoiesis regulating, tissue growth, activin/inhibin,
CC chemotactic/chemokinetic, haemostatic and thrombolytic,
CC receptor/ligand, anti-inflammatory or tumour inhibition activities.
CC It can also be used to research or treat diseases/disorders related
CC to its function.
CC The partial cDNA clone AP162 was 1st isolated from a human adult
CC placenta cDNA library. The partial cDNA clones AM931, AM610, AM340,
CC AM282, AK647, AK583, AK533 and AK296 were 1st isolated from a human
CC foetal kidney cDNA library. The partial cDNA clones H617 and B89
CC were 1st isolated from a human peripheral blood monocyte cell (Tnl
CC or Tl2) cDNA library. The partial cDNA clone AM191 was 1st isolated
CC from a human ovary (PA-1 teratocarcinoma) cDNA library. The partial
CC cDNA clones AT711, AT705 and AT319 were 1st isolated from a human
CC lymphocyte and dendritic cell cDNA library. The partial cDNA clones
CC AS34 and AS32 were 1st isolated from a human foetal brain cDNA
CC library. The partial cDNA clone AR260 was 1st isolated from a human
CC adult retina cDNA library. The partial cDNA clones K640 and K39
CC were 1st isolated from a murine bone marrow (stromal cell line
CC FCM-4) cDNA library.
SQ Sequence 159 AA;
```

```
Query Match 10.4%; Score 195.5; DB 1; Length 159;
Best Local Similarity 43.4%; Pred. No. 2,3e-12;
Matches 36; Conservative 21; Mismatches 25; Indels 1; Gaps 1;
```

```
QY 3 PONTSPCLLLLVIGAVIGRDFYKILGVPRASIKDIRKAYRKLALQLHPRNPPDPQ 62
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 4 POSIFPAICILMITELILASKSYDILGVPRKASERQIKKAFHKLAMKYPDKN-KSPD 62
```

```
QY 63 AOEKFDUGAAYEVLSDSEKRRQ 85
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 63 AEAKFREIAEAYETLSDANXTKR 85
```

```
RESULT 15
ID W59132 standard; Protein; 438 AA.
AC W59132;
DT 11-SEP-1998 (first entry)
DE Mus musculus T1b Interactor (MTI-3) protein.
KW serine protease; tub interactor; treatment; obesity; cachexia;
KW anorexia nervosa; diabetes; cell cycle progression; apoptosis;
KW neurodegenerative disease; Alzheimer's disease; drug screening;
KW Parkinson's disease; Huntington's chorea; detection; diagnosis;
KW amyotrophic lateral sclerosis; spinocerebellar degeneration.
OS Mus musculus.
PN W09812302-A1.
PD 26-MAR-1998.
PF 05-SEP-1997; U15627.
PR 21-JUL-1997; US-897340.
PR 17-SEP-1996; US-715032.
PA (MILL-) MILLENNIUM PHARM INC.
PI Errata PR, Glimco CJ;
DR WPI: 98-217246/19.
DR N-PSDB: V11858.
PT Tub interactor genes - used to develop products for the treatment
PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
PT diabetes
PS Claim 28; Fig 4; 120pp; English.
CC The sequence is that encoding the T1b Interactor protein (MTI-3)
CC which is a putative serine protease. T1 genes function
CC in biochemical pathways involved in weight control and
CC related disorders. The products can be used for treating
CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
CC or a related disorder such as diabetes. The products can
CC also be used to modulate cell cycle progression and apoptosis.
CC They can be used for treating neurodegenerative diseases
CC which are characterised by apoptosis, including Alzheimer's
CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
CC lateral sclerosis or spinocerebellar degenerations. The
CC products can also be used for detection, diagnosis and
CC drug screening.
SQ Sequence 438 AA;
```

```
Query Match 10.3%; Score 194; DB 1; Length 438;
Best Local Similarity 38.8%; Pred. No. 1,4e-11;
Matches 45; Conservative 23; Mismatches 26; Indels 22; Gaps 4;
```

```
QY 24 RDEYKILGVPRASIKDIRKAYRKLALQLHPRNPD----DPQAEKFDUGAAYEVL 78
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 324 KDYKILGVPRASIKDIRKAYRKLALHPRNPDHRSASAEVQEKKEKFEVGEAFILS 383
QY 79 DSEKRYQYDT---YGEGLADGHOSHCDFISHF-----GDGFMFG 118
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 384 DPKKRTYDSQDLDEGMNMGDFDAN-NIFKAFGGGGRSFEASGPNFYFQFG 438
```

Search completed: June 30, 2000, 09:57:22
Job time: 2723 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 11:53:59 ; Search time 49.41 Seconds
(without alignments)
3619.903 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376
Sequence: 1 TCTCACCAGGACTCGGACT.....TTTATTTCATATGCAGT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :
1: /cgn2_6/prodata/2/lna/5A_COMB.seq:*
2: /cgn2_6/prodata/2/lna/5B_COMB.seq:*
3: /cgn2_6/prodata/2/lna/5C_COMB.seq:*
4: /cgn2_6/prodata/2/lna/5D_COMB.seq:*
5: /cgn2_6/prodata/2/lna/6_COMB.seq:*
6: /cgn2_6/prodata/2/lna/PCRUS_COMB.seq:*
7: /cgn2_6/prodata/2/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1374	99.9	1376	4	US-08-868-288A-2
2	1374	99.9	1376	5	US-09-235-373-2
3	1374	99.9	1376	5	US-09-388-993-2
4	91.4	6.6	1330	4	US-08-868-288A-4
5	91.4	6.6	1330	5	US-09-235-373-4
6	91.4	6.6	1330	5	US-09-388-993-4
7	65.2	4.7	2349	4	US-08-974-546-2
8	65	4.7	672	2	US-08-486-955A-6
9	59.4	4.3	7218	1	US-08-232-653-14
10	53.4	3.9	1515	1	US-08-221-816B-1
11	53.4	3.9	1687	1	US-08-143-219-26
12	48	3.5	1700	4	US-08-897-340-4
13	47.8	3.5	4320	4	US-08-472-534-4
14	44.8	3.3	1756	4	US-08-879-260-3
15	40.8	3.0	20235	2	US-07-642-73AC-3
16	39.6	2.9	3748	3	US-08-686-417-2
17	38.8	2.8	1860	6	US-08-531-644-3
18	38.8	2.8	1860	6	PCT-US93-04102-3
19	38.6	2.8	1055	6	PCT-US93-04102-3
20	38.6	2.8	1055	6	PCT-US95-04896-1
21	38.6	2.8	5852	1	US-07-867-106-2
22	38	2.8	2277	2	US-08-676-967-2
23	38	2.8	2277	2	US-08-676-974-2
24	38	2.8	2277	2	US-09-098-487-2
25	36.8	2.7	2064	1	US-08-343-428-1
26	36.8	2.7	22846	4	US-08-469-461-3
27	36.8	2.7	22846	5	US-07-890-609-3

C	28	36.6	2.7	1164	7	5240849-4	Patent No. 5240849
C	29	36.6	2.7	2749	7	5240849-1	Patent No. 5240849
	30	35	2.5	2326	3	US-08-231-193A-41	Sequence 41, Appl
	31	35	2.5	2326	4	US-08-486-273A-41	Sequence 41, Appl
	32	35	2.5	2326	5	US-08-480-474-41	Sequence 41, Appl
	33	35	2.5	3243	3	US-08-231-193A-44	Sequence 44, Appl
	34	35	2.5	3243	4	US-08-486-273A-44	Sequence 44, Appl
	35	35	2.5	3243	5	US-08-480-474-44	Sequence 44, Appl
	36	35	2.5	3698	3	US-08-231-193A-43	Sequence 43, Appl
	37	35	2.5	3698	4	US-08-486-273A-43	Sequence 43, Appl
	38	35	2.5	3698	5	US-08-480-474-43	Sequence 43, Appl
	39	35	2.5	4002	3	US-08-231-193A-53	Sequence 53, Appl
	40	35	2.5	4002	4	US-08-486-273A-53	Sequence 53, Appl
	41	35	2.5	4002	5	US-08-480-474-53	Sequence 53, Appl
	42	35	2.5	4017	3	US-08-231-193A-49	Sequence 49, Appl
	43	35	2.5	4017	4	US-08-486-273A-49	Sequence 49, Appl
	44	35	2.5	4017	5	US-08-480-474-49	Sequence 49, Appl
	45	35	2.5	4053	3	US-08-231-193A-47	Sequence 47, Appl

ALIGNMENTS

RESULT 1
US-08-868-288A-2
Sequence 2, Application US/0868288A
Patent No. 5922567
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
OPERATING SYSTEM: IBM Compatible
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868, 288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1376 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: SYNORAB01
CLONE: 136466
US-08-868-288A-2

Query Match 99.9%; Score 1374; DB 4; Length 1376;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Gaps 0;

QY 1 TCTACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGCTAGTAGTG 60
 DB 1 TCTACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGCTAGTAGTG 60
 QY 61 TCTCTGCGGAGCAGGAGAAACCCCGCCCGCGGTGTAGGGCGGCTCACAGGGCCGG 120
 DB 61 TCTCTGCGGAGCAGGAGAAACCCCGCCCGCGGTGTAGGGCGGCTCACAGGGCCGG 120
 QY 121 GTGGGCTGGAGCGAGCGGCGGAGGAGGCTGTGAGAGGTGTGGAACAGAGACC 180
 DB 121 GTGGGCTGGAGCGAGCGGCGGAGGAGGCTGTGAGAGGTGTGGAACAGAGACC 180
 QY 181 CGGAGACAGAGAACCCAGGCTCCGAGAACCTGAGCACCTTTGGCTGTGCTCTATAC 240
 DB 181 CGGAGACAGAGAACCCAGGCTCCGAGAACCTGAGCACCTTTGGCTGTGCTCTATAC 240
 QY 241 CTCAATCGGGGGGTGATGTCGGGAGAGATTTCTATAGATCTTGGGGGTGCTCGAAGT 300
 DB 241 CTCAATCGGGGGGTGATGTCGGGAGAGATTTCTATAGATCTTGGGGGTGCTCGAAGT 300
 QY 301 GCCTCTATAAGATATTTAAAAAGCCCTATAGAACTAGCCCTGACGCTTCATCCGAGC 360
 DB 301 GCCTCTATAAGATATTTAAAAAGCCCTATAGAACTAGCCCTGACGCTTCATCCGAGC 360
 QY 361 CGGAACCTGATGATCCACAAGCCAGAGAAATTCAGGATCTGGTCTTATAGAG 420
 DB 361 CGGAACCTGATGATCCACAAGCCAGAGAAATTCAGGATCTGGTCTTATAGAG 420
 QY 421 GTTCTGATAGTATGAGAAACGGAACAGACATCTTATGTTGAAGAGATTAAAA 480
 DB 421 GTTCTGATAGTATGAGAAACGGAACAGACATCTTATGTTGAAGAGATTAAAA 480
 QY 481 GATGGTCATAGAGCTCCCATGAGACATTTTTCACACTTCTTGGGATTTTGGTTTC 540
 DB 481 GATGGTCATAGAGCTCCCATGAGACATTTTTCACACTTCTTGGGATTTTGGTTTC 540
 QY 541 ATGTTTGGAGAGAACCCCTGTCAGCAAGACAAATATTCGAAGAGAGATGATATTAT 600
 DB 541 ATGTTTGGAGAGAACCCCTGTCAGCAAGACAAATATTCGAAGAGAGATGATATTAT 600
 QY 601 GTAGATCTAGAGCTTGTGGAAGATATGAGCAAAATTTTGTGGAAGATGATAGA 660
 DB 601 GTAGATCTAGAGCTTGTGGAAGATATGAGCAAAATTTTGTGGAAGATGATAGA 660
 QY 661 AACAAACCTGTGGCAAGGAGGCTCTGGCAACGGAAGTGTGCGCAAGAGATG 720
 DB 661 AACAAACCTGTGGCAAGGAGGCTCTGGCAACGGAAGTGTGCGCAAGAGATG 720
 QY 721 CGGACCAACGAGTGGGCGCTGGGCGCTTCCAAATGACCCAGAGGTGTGCGAGAA 780
 DB 721 CGGACCAACGAGTGGGCGCTGGGCGCTTCCAAATGACCCAGAGGTGTGCGAGAA 780
 QY 781 TGCCCTAATGTCAAACTAGTGAAGAAAGCAAGCTGGAATAGAAAATAGGCGCTGG 840
 DB 781 TGCCCTAATGTCAAACTAGTGAAGAAAGCAAGCTGGAATAGAAAATAGGCGCTGG 840
 QY 841 GTGAGAGACGCGATGAGTACCCCTTTATTTGGAAGGTGAGCCTGATGAGGGAG 900
 DB 841 GTGAGAGACGCGATGAGTACCCCTTTATTTGGAAGGTGAGCCTGATGAGGGAG 900
 QY 901 CCGGAGATTTACGGTTCGGAATCAAGTGTCAAGCACCATATTTTAAAGGAGAGCA 960
 DB 901 CCGGAGATTTACGGTTCGGAATCAAGTGTCAAGCACCATATTTTAAAGGAGAGCA 960
 QY 961 GATGATTTGTACAAATGTGACAGTCTATAGTTGATGATGATGATGATGATGATG 1020
 DB 961 GATGATTTGTACAAATGTGACAGTCTATAGTTGATGATGATGATGATGATGATG 1020
 QY 1021 GATATTTACTCACTTGGATGTGACAAAGGTATATTTCCCGGAGTAAAGTACCAAGGCCA 1080
 DB 1021 GATATTTACTCACTTGGATGTGACAAAGGTATATTTCCCGGAGTAAAGTACCAAGGCCA 1080

QY 1081 GGAGCGAANTANTGGAAGAAAGGGGCTCCCAACTTTGACAAACAATATCAAG 1140
 DB 1081 GGAGCGAANTANTGGAAGAAAGGGGCTCCCAACTTTGACAAACAATATCAAG 1140
 QY 1141 GGCTCTTTGATATACATTTTGTGATGTGATTTTCCAAAGACAGTTTAAACAGAGGAG 1200
 DB 1141 GGCTCTTTGATATACATTTTGTGATGTGATTTTCCAAAGACAGTTTAAACAGAGGAG 1200
 QY 1201 AGAGAGATCAAAACAGTACTGAAAACAAGGGTCACTGAGAGATATCAATGAGACTG 1260
 DB 1201 AGAGAGATCAAAACAGTACTGAAAACAAGGGTCACTGAGAGATATCAATGAGACTG 1260
 QY 1261 CAAGATATTGAGAGTGAATATAATGAGCTTTGTTTAAATAGGATATACAGATATT 1320
 DB 1261 CAAGATATTGAGAGTGAATATAATGAGCTTTGTTTAAATAGGATATACAGATATT 1320
 QY 1321 TATTATCTGCAAGGTTTTTTTGTGTGTGTTTTGTTTTATTTTCAATATGCAAGT 1376
 DB 1321 TATTATCTGCAAGGTTTTTTTGTGTGTGTTTTGTTTTATTTTCAATATGCAAGT 1376

RESULT 2
 US-09-235-373-2
 ; Sequence 2, Application US/09235373
 ; Patent No. 6001598
 ; GENERAL INFORMATION:
 ; APPLICANT: Au-Young, Janice
 ; APPLICANT: Lai, Preeti
 ; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
 ; STREET: 3174 Porter Drive
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94304
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FASTSEQ for Windows Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/235,373
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/868,288
 ; FILING DATE: June 3, 1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Billings, Lucy J.
 ; REGISTRATION NUMBER: 36,749
 ; REFERENCE/DOCKET NUMBER: PF-0309 US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 415-855-0555
 ; TELEFAX: 415-845-4166
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1376 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; IMMEDIATE SOURCE:
 ; LIBRARY: SYNORAB01
 ; CLONE: 136466
 ; US-09-235-373-2

Query Match 99.9%; Score 1374; DB 5; Length 1376;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TCTACCGGAGCTCGGAGCTCCCGGAAAGTGAGACCGGAGAGAGGGGCTAGTAGTG 60

|||||
Db 1 TCTACCGGGGACTCGGGACTCCGGGAAGTGGACCCGGCAGAGAGGGGGCTAGCTACTG 60
QY 61 TCTCTGGGACACGAGGAGACCCCGCGCCCCCGGTGTGAGGGGCTCCACAGGGCGG 120
Db 61 TCTCTGGGACACGAGGAGACCCCGCGCCCCCGGTGTGAGGGGCTCCACAGGGCGG 120
QY 121 GTGGGCTGGGAGCGGACCGGGCGGAGAGGCTGTGAGGATGTGTGGAACAGAGCC 180
Db 121 GTGGGCTGGGAGCGGACCGGGCGGAGAGGCTGTGAGGATGTGTGGAACAGAGCC 180
QY 181 CGGGACAGAGAAACCATGGCTCCGACAGACCTGTTTCTGTTTCTGTTATAC 240
Db 181 CGGGACAGAGAAACCATGGCTCCCGACAGACCTGAGACCTTTTCTGTTTCTGTTATAC 240
QY 241 CTCATCGGGCGGTGATGTCGGGACGAGATTCTATAGATCTTGGGGGTGCTCGAAGT 300
Db 241 CTCATCGGGCGGTGATGTCGGGACGAGATTCTATAGATCTTGGGGGTGCTCGAAGT 300
QY 301 GCCCTATTAAGATATTAAGGCTATATGAAACTAGCCCTGACAGCTTCATCCGAC 360
Db 301 GCCCTATTAAGATATTAAGGCTATATGAAACTAGCCCTGACAGCTTCATCCGAC 360
QY 361 CGGAACCTGATGATCCACAAGCCGAGAGAAATCCAGATCTGGGTGCTGTTATGAG 420
Db 361 CGGAACCTGATGATCCACAAGCCGAGAGAAATCCAGATCTGGGTGCTGTTATGAG 420
QY 421 GTTCTGTCAATAGTGAAGAAACGGAACAGTACGATCTTATGTTGGAAGATTAAGA 480
Db 421 GTTCTGTCAATAGTGAAGAAACGGAACAGTACGATCTTATGTTGGAAGATTAAGA 480
QY 481 GATGCTATAGAGCTCCCATGAGACATTTTTCACACTCTTGGGGATTTGGTTTC 540
Db 481 GATGCTATAGAGCTCCCATGAGACATTTTTCACACTCTTGGGGATTTGGTTTC 540
QY 541 ATGTTTGAGAGAACCCCTGTCACAGACAGAAATATCCAGAGAGATGATATT 600
Db 541 ATGTTTGAGAGAACCCCTGTCACAGACAGAAATATCCAGAGAGATGATATT 600
QY 601 GTAGATCTAGAACTCTTTGGAGAAATATATGAGAAATTTTGGGAAGTATAGA 660
Db 601 GTAGATCTAGAACTCTTTGGAGAAATATATGAGAAATTTTGGGAAGTATAGA 660
QY 661 AACAACTGTGGGAGGAGGCTCTGGCAAGGAGGAGTGTGCTGGCAGAGATG 720
Db 661 AACAACTGTGGGAGGAGGCTCTGGCAAGGAGGAGTGTGCTGGCAGAGATG 720
QY 721 CGGACCAACAGCTGGGGCCCTGGGCGCTTCCAAATGACCCGAGAGGTGTGCGACGAA 780
Db 721 CGGACCAACAGCTGGGGCCCTGGGCGCTTCCAAATGACCCGAGAGGTGTGCGACGAA 780
QY 781 TGCCCTAATGTCAAACTAGTGAATGAAGAAAGGCTGGAATGAAATAGACCTGGG 840
Db 781 TGCCCTAATGTCAAACTAGTGAATGAAGAAAGGCTGGAATGAAATAGACCTGGG 840
QY 841 GTGAGAACGCGCATGAGTACCCCTTATTTGAGAGAGTGAAGCTCACGTGAGTGGGAG 900
Db 841 GTGAGAACGCGCATGAGTACCCCTTATTTGAGAGAGTGAAGCTCACGTGAGTGGGAG 900
QY 901 CCTGAGATTTAGGTTCCGAATCAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGGA 960
Db 901 CCTGAGATTTAGGTTCCGAATCAAGTTGTCAAGCACCCCAATATTTGAAAGAGAGGA 960
QY 961 GATGATTTGTACACAAATGTGACAGTCTCATAGTTAGTCACTGTTGGCTTTGAGATG 1020
Db 961 GATGATTTGTACACAAATGTGACAGTCTCATAGTTAGTCACTGTTGGCTTTGAGATG 1020
QY 1021 GATTTTACTACTGTGATGGTCAAGGTACATTTTCCGGGATTAAGATACACAGGCCA 1080
Db 1021 GATTTTACTACTGTGATGGTCAAGGTACATTTTCCGGGATTAAGATACACAGGCCA 1080
QY 1081 GGACGGAANTGGAAGAAAGGGAGGGCTCCCAACTTTGACAAACAATATCAAG 1140
|||||

Db 1081 GGACGGAANTGGAAGAAAGGGAGGGCTCCCAACTTTGACAAACAATATCAAG 1140
QY 1141 GGCTCTTTGATTAATCACTTTTGTATGTGATTTTCCAAAAGACGTTTACAGAGAGCG 1200
Db 1141 GGCTCTTTGATTAATCACTTTTGTATGTGATTTTCCAAAAGACGTTTACAGAGAGCG 1200
QY 1201 AGAAGAGTATCAAAAGCTCTGAAACAAAGGTGACGCAAGAGTATACATGAGCTG 1260
Db 1201 AGAAGAGTATCAAAAGCTCTGAAACAAAGGTGACGCAAGAGTATACATGAGCTG 1260
QY 1261 CAAGATATTTAGAGTGAATTAATTTGACTTTGTTTAAATAGTAAATAGCGATATT 1320
Db 1261 CAAGATATTTAGAGTGAATTAATTTGACTTTGTTTAAATAGTAAATAGCGATATT 1320
QY 1321 TATTAATCGAAGGTTTTTTGTGTGTTTGTGTTTATTTTCAATATGCAAGT 1376
Db 1321 TATTAATCGAAGGTTTTTTGTGTGTTTGTGTTTATTTTCAATATGCAAGT 1376

RESULT 3
US-09-388-993-2
; Sequence 2, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
; US-09-388-993-2

Query Match 99.9%; Score 1374; DB 5; Length 1376;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCTACCGGGGACTCGGGACTCCGGGAAGTGGACCGGCAAGAGGGGGCTAGCTAGCTG 60
Db 1 TCTACCGGGGACTCGGGACTCCGGGAAGTGGACCGGCAAGAGGGGGCTAGCTAGCTG 60

```

QY 61 TCTCTGCGGACCAGGAGAGACCCCGGCGCCCGGCGTGTGAGGCGGCGCTCACAGGGCGG 120
DB 61 TCTCTGCGGACCAGGAGAGACCCCGGCGCCCGGCGTGTGAGGCGGCGCTCACAGGGCGG 120
QY 121 GTGGGCTGGCGAGCCGAGCGGCGGCGAGAGGCTGTGAGAGTGTGTGAAACAGAGAC 180
DB 121 GTGGGCTGGCGAGCCGAGCGGCGGCGAGAGGCTGTGAGAGTGTGTGAAACAGAGAC 180
QY 181 CGGACAGAGAAACCATGGCTCCGAGAACTGACACCTTTGGCTGTGTCTATAC 240
DB 181 CGGACAGAGAAACCATGGCTCCGAGAACTGACACCTTTGGCTGTGTCTATAC 240
QY 241 CTCAATCGGGGCGGTGATTCGCGAGAGATTTCTATAGATCTTGGGGTCTCGAAGT 300
DB 241 CTCAATCGGGGCGGTGATTCGCGAGAGATTTCTATAGATCTTGGGGTCTCGAAGT 300
QY 301 GCGCTTATAAAGATTTAAAAAGGCGCTATAGAAACTAGCCCTGACCTTCATCCGAC 360
DB 301 GCGCTTATAAAGATTTAAAAAGGCGCTATAGAAACTAGCCCTGACCTTCATCCGAC 360
QY 361 CGGAACCTCGATGATCCAGAGCCGAGAAATTCAGATCTGGGTGCTGCTTATGAG 420
DB 361 CGGAACCTCGATGATCCAGAGCCGAGAAATTCAGATCTGGGTGCTGCTTATGAG 420
QY 421 GTTCTGATAGTGTAGAGAAACGAAACAGTACATCTTATGCTGAAGAAGATTAAA 480
DB 421 GTTCTGATAGTGTAGAGAAACGAAACAGTACATCTTATGCTGAAGAAGATTAAA 480
QY 481 GATGGTATCAGAGCTCCAGAGAGACATTTTTCACACTTTTGGGGATTTGGTTTC 540
DB 481 GATGGTATCAGAGCTCCAGAGAGACATTTTTCACACTTTTGGGGATTTGGTTTC 540
QY 541 ATGTTTGAGAGAACCCCTCGTCAGCAGACAGAAATATTCAGAGAGAGATATTATT 600
DB 541 ATGTTTGAGAGAACCCCTCGTCAGCAGACAGAAATATTCAGAGAGAGATATTATT 600
QY 601 GTATATCTAGAAGTCACTTTTGGAAAGATATGACAGAAATTTTGTGGAAGTATTAGA 660
DB 601 GTATATCTAGAAGTCACTTTTGGAAAGATATGACAGAAATTTTGTGGAAGTATTAGA 660
QY 661 AACAAACCTGTGGCAAGGAGGCTCCTGGCAACGGAAGGCAATTCGCGCAAGAGATG 720
DB 661 AACAAACCTGTGGCAAGGAGGCTCCTGGCAACGGAAGGCAATTCGCGCAAGAGATG 720
QY 721 CGGACCAACCAAGCTGGGCGCTTCCAAATGACCCAGAGAGTGTCTCGACGAA 780
DB 721 CGGACCAACCAAGCTGGGCGCTTCCAAATGACCCAGAGAGTGTCTCGACGAA 780
QY 781 TGGCCTTAATGTCAAACTAGTGAATGAAGACGAACGCTGGAATAGAAATGAGCCTGG 840
DB 781 TGGCCTTAATGTCAAACTAGTGAATGAAGACGAACGCTGGAATAGAAATGAGCCTGG 840
QY 841 GTGAGAGACGCGATGAGTACCCCTTATGGAAGAGTACGCTACGATGGGGAG 900
DB 841 GTGAGAGACGCGATGAGTACCCCTTATGGAAGAGTACGCTACGATGGGGAG 900
QY 901 CCGAGAGATTTACGGTCCGAATCAAAAGTTGTCAGACCCCAATATTGGAAGAGAGA 960
DB 901 CCGAGAGATTTACGGTCCGAATCAAAAGTTGTCAGACCCCAATATTGGAAGAGAGA 960
QY 961 GATGATTTGTACCAATGTGACAGTCTATTAAGTGAAGTACGCTGTGGCTTTGAGATG 1020
DB 961 GATGATTTGTACCAATGTGACAGTCTATTAAGTGAAGTACGCTGTGGCTTTGAGATG 1020
QY 1021 GATATTACTACTTGGATGTCAACAAGTACATATTTCCGGGATTAAGTACACAGGCA 1080
DB 1021 GATATTACTACTTGGATGTCAACAAGTACATATTTCCGGGATTAAGTACACAGGCA 1080
QY 1081 GGAGGGAATTAATGGAAGAAAGGGAGGCTCCCACTTTGACAAACAATATCAAG 1140
DB 1081 GGAGGGAATTAATGGAAGAAAGGGAGGCTCCCACTTTGACAAACAATATCAAG 1140

```

```

QY 1141 GCGCTTTGATTAATCACTTTTGTATGTGATTTCCAAAAGACGTTAAACAGAGAACGC 1200
DB 1141 GCGCTTTGATTAATCACTTTTGTATGTGATTTCCAAAAGACGTTAAACAGAGAACGC 1200
QY 1201 AGAAGAGTATCAAAACAGCTCTGAAACAGGGTCAAGAGATATCAATGAGACTG 1260
DB 1201 AGAAGAGTATCAAAACAGCTCTGAAACAGGGTCAAGAGATATCAATGAGACTG 1260
QY 1261 CAAGGATTTGAGAGTGAATTAATTTGAGCTTTGTTTAAATAAGTAAATACGATATT 1320
DB 1261 CAAGGATTTGAGAGTGAATTAATTTGAGCTTTGTTTAAATAAGTAAATACGATATT 1320
QY 1321 TATATCTGAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1376
DB 1321 TATATCTGAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1376

RESULT 4
US-08-868-288A-4
; Sequence 4: Application US/08868288A
; Patent No. 592567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAL-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; City: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868, 288A
; FILING DATE: June 3, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1350 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNTZRA01
; CLONE: 260873
; US-08-868-288A-4

```

```

Query Match 6.6%; Score 91.4; DB 4; Length 1330;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;
QY 268 GATTTCTATTAAGATCTTGGGGTGGCTCGAAGTGCCTATATAAAGATATAAAGGCC 327
DB 163 GATTAATCTGAAGTCTTGAAGGCTGCAAGACATGCTTACCCGAGATATTAAGAGGCA 222
QY 328 TATAGAAACTAGCCCTGACCTTCATCCGAGCCGGAACCTG--ATGATCCACAAGCC 384

```

Db 223 TATCGGAACCTGGCAGAGTGGCATCCAGATAAAAATCTGTGAGATTAAGAAAGCA 282
QY 385 CAGAGAAATTCAGAGATCTGGGTGCTTATGAGTCTGTGACATAGTGAAGAACG 444
Db 283 GAGAGAAATTCAGAGATGAGGAGGATATGAGTCTGTGAGTGAAGAACG 342
QY 445 AAACAGTACGATCTATGAGTGAAGAGATTAAGATG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGTTG 383

RESULT 5

US-09-235-373-4
; Sequence 4, Application US/09235373
; Patent No. 6001398
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMT2RAT01
; CLONE: 260873
US-09-235-373-4

Query Match 6.6%; Score 91.4; DB 5; Length 1330;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATTAAGATCTTGGGGGTGCTCGAAGTGCCTCTATAAGATATTAAGAGCC 327
Db 163 GATTACTATGAAGTCTAGGCGTGCAGAGATGCTCCACCGAGATATTAAGAGCA 222
QY 328 TATAGAAATTCAGACCTGCGAGCTTCATCCGACCGGAAACCTG---ATGATCCAAAGCC 384
Db 223 TATCGGAACCTGGCAGTGAAGTGCATCCAGATTAAGATCTGTGATTAAGAGCA 282
QY 385 CAGAGAAATTCAGAGTCTGGGTGCTTATGAGTCTGTGACATAGTGAAGAACG 444
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGTTG 383

Db 283 GAGAGAAATTCAGAGATGAGGAGGATATGAGTGTGCGATGCTGAAGAACG 342
QY 445 AAACAGTACGATCTATGAGTGAAGAGATTAAGATG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGTTG 383

RESULT 6

US-09-388-993-4
; Sequence 4, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAM-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-845-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HMT2RAT01
; CLONE: 260873
US-09-388-993-4

Query Match 6.6%; Score 91.4; DB 5; Length 1330;
Best Local Similarity 66.5%; Pred. No. 3.8e-17;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

QY 268 GATTCTATTAAGATCTTGGGGGTGCTCGAAGTGCCTCTATAAGATATTAAGAGCC 327
Db 163 GATTACTATGAAGTCTAGGCGTGCAGAGATGCTCCACCGAGATATTAAGAGCA 222
QY 328 TATAGAAATTCAGACCTGCGAGCTTCATCCGACCGGAAACCTG---ATGATCCAAAGCC 384
Db 223 TATCGGAACCTGGCAGTGAAGTGCATCCAGATTAAGATCTGTGATTAAGAGCA 282
QY 385 CAGAGAAATTCAGAGTCTGGGTGCTTATGAGTCTGTGACATAGTGAAGAACG 444
Db 283 GAGAGAAATTCAGAGATGAGGAGGATATGAGTCTGTGATTAAGAGCA 342
QY 445 AAACAGTACGATCTATGAGTGAAGAGATTAAGATG 485
Db 343 GACATCTATGACAAATATGCGCAAGAGATTAATGTTG 383

RESULT 7

US-08-974-546-2
; Sequence 2, Application US/08974546
; Patent No. 5945287
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preethi
; APPLICANT: Corley, Neil C.
; APPLICANT: Shah, Purni
; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/974,546
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0428
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2349 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: BRAITVT21
; CLONE: 2525691
; US-08-974-546-2

Query Match 4.7%; Score 65.2; DB 4; Length 2349;
Best Local Similarity 56.3%; Pred. No. 2.6e-09;

Matches 143; Conservative 0; Mismatches 108; Indels 3; Gaps 1;

OY 262 GGACGAGATTCTTAAGATCTGGGGGTCCTCGAAGCCCTCTTAAGGATTATAA 321
DB 105 GGAAGAGATTATTAAGATCTGGGATCCCATCGGGGCAAGCAGAGATGATCAAG 164
OY 322 AAGGCTATAGGAACCTGACCTTCATCCCGACCGGAAACCTGATGATCAACA 381
DB 165 AAGGCTACCGGAAGATGGCTTGAAGTACCAACCCGACGAAGA---TAAAGACCCCAAC 221
OY 382 GCCCAGGAATAATTCAGATCTGGGTGCTGCTTATGAGCTTCTGATAGTAGAGAAA 441
DB 222 GCTGAGGAGAAGTTAAAGAGATTGCGAGGCTTATGATGCTAAGTAGCAACCAAGAAA 281
OY 442 CGAAGACAGTACGATCTATGATGTAAGAGATTAAAGATGTCATCAGAGCTCCAT 501
DB 282 CGGGGCTTATATGACCAATGATGGGAGAGAGCCCTGAAGACCCGGGGTGGCATCAGGT 341
OY 502 GGAGACATTTTTC 515
DB 342 GGCTCCAGTGGCTC 355

RESULT 8

US-08-486-955A-6
; Sequence 6, Application US/08486955A
; Patent No. 5747299
; GENERAL INFORMATION:
; APPLICANT: FATHMAN, Garrison
; APPLICANT: BLOOM, Debra
; TITLE OF INVENTION: Anergy Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hombach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,955A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Berttram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A59741-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-486-955A-6

Query Match 4.7%; Score 65; DB 2; Length 672;
Best Local Similarity 56.9%; Pred. No. 1.3e-09;

Matches 119; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

OY 256 ATTCGCGAGAGATTCTTAAGATCTGGGGTGGCTCGAAGGCTCTATAAAGAT 315
DB 31 ACTTCGGGGAGATCGTTATACCATGTTCTTGACCTGGACAAAGATGCAACCTCAGATGAC 90
OY 316 ATTAAGAGGCTTAAGAACTACCCCTGCAGCTTCATCCGACCGGAACCTGATGAT 375
DB 91 ATTAAGAGGCTTAAGAACTACCCCTGCAGCTTCATCCGACCGGAACCTGATGAT 150
OY 376 CCAGAGCCGAGGAATTCAGATCTGGGTGCTGCTTATGAGTCTCTGATAGT 435
DB 151 CCAGAGGCTGCAGCAAGATTTAAGAGATTAAACAACGACACGCCATCTTGACAGAGCC 210
OY 436 GAGAAACGGAACAGTACGATCTATGAG 464
DB 211 ACGAAAGAAACATTTATGACAGATATG 239

RESULT 9

US-08-232-463-14/C
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHIEFLINGER, F.
; APPLICANT: FALKNER, F. G.

	Query Match Similarity	4.3%;	Score 59.4;	DB 1;	Length 7218;	
	Pest Local Similarity	7.6%;	Pred. No. 2, 6e-07;			
	Matches	33;	Conservative	224;	Mismatches 180;	Indels 0; Gaps 0;
OY	534	TTGGGATTTTGGTTTCATGTTTGAGGAACCCCTGTCAGCAGACAAGAATATTCCAA	583			
		: : : : : : : :				
Db	1450	TAGAGAATTTTGCACRRR	1391			
OY	584	GAGCAAGTCATTTTGTGTAGATCTAGAAGTCACTTTGSAABAAGTATATGCAGCAAAT	643			
Db	1390	RRR	1331			
OY	644	TTGTGGAAGTAGTWTAGAAAACAACCTGTGTGCAAGCAGCTCCTGCCAACGAAAGTCA	703			
Db	1330	RRR	1271			
OY	704	ATTGTGGCACAGATGCGGACCACCCAGCTGGGCCCTTGCCAAATGACCACAG	763			
Db	1270	RRR	1211			
OY	764	AGGTGTGTGCACAGATGCCCTAATGTCAACTGTGAATAAGAACAGCAGCTGGAG	823			
Db	1210	RRR	1151			
OY	834	TAGAAATAGAGCCTGGGCTGAGAGACGCGCATGAGTAGTACCCTTTATTGAGAGAGTGAC	883			
Db	1150	RRR	1091			
OY	884	CTCAGCTGATGGGAGCCTGGAGATTTTCGGTCCGAATCAAAGTTGCAAGCACCCA	943			
Db	1090	RRRRRRRRRRRRRRRRRRRRRRRRRAATCGCAACCTCCCTGACCTGAGCAAGCTGTGAA	1031			

[illegible]

STREET: Abbott Laboratories D377/AP6D-2 One Abbott
STREET: Park Rd
CITY: Abbott Park
STATE: IL
COUNTRY: US
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/642,734C
FILING DATE: 17-JAN-91
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Danckers, Andreas M
REGISTRATION NUMBER: 32652
REFERENCE/DOCKET NUMBER: 4952.US.01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708-937-9396
TELEFAX: 708-938-2623
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 20235 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Saccharopolyspora erythraea
STRAIN: NRRL 238
FEATURE:
NAME/KEY: CDS
LOCATION: 19..10722
OTHER INFORMATION: /codon_start=19
OTHER INFORMATION: /function="gene eryA"
OTHER INFORMATION: /product="eryA ORF2 encoding modules 3 & 4 for
OTHER INFORMATION: 6-deoxyerythronolide B"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 19..4470
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 97..1482
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacyl ACP synthase of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1693..2670
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 3406..3921
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4171..4428
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 3"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 4471..10722
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921

LOCATION: 4471..5847
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 6054..7026
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 7165..9216
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: dehydratase and enoylreductase domains m"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 9433..9984
OTHER INFORMATION: /function="approximate span
OTHER INFORMATION: beta-ketoreductase of module 4"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10225..10483
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 4"
FEATURE:
NAME/KEY: CDS
LOCATION: 10723..20235
OTHER INFORMATION: /codon_start=10723
OTHER INFORMATION: /function="gene eryA"
OTHER INFORMATION: /product="orf3 encoding modules 5 & 6
OTHER INFORMATION: 6-deoxyerythronolide B formatio"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10723..15165
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 10831..12174
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 12379..13350
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14062..14610
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoreductase of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 14857..15114
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyl carrier domain of module 5"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15166..20235
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 15172..16569
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: beta-ketoacylACPSynthase domain of modul"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 16768..17721
OTHER INFORMATION: /function="approximate span of
OTHER INFORMATION: acyltransferase domain of module 6"
FEATURE:
NAME/KEY: misc_feature
LOCATION: 18379..18921

```
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: beta-ketoreductase domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19149..19398
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: acyl carrier domain of module 6"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19492..20235
; OTHER INFORMATION: /function="approximate span of
; OTHER INFORMATION: thioesterase domain of module 6"
; US-07-642-734C-3
```

```
Query Match 3.0%; Score 40.8; DB 2; Length 20235;
Best Local Similarity 54.7%; Pred. No. 0.14;
Matches 81; Conservative 0; Mismatches 67; Indels 0; Caps 0;

QY 70 ACCAGGAGAGACCCCGCGCGCCCGCGGTGTGAGGCGGCGCTCACAGGGCCGCGTGGCTGG 129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10933 ATGCGGAGAGCCCGCGCGCGCGCGCTGGAGCGGAGCCCGCGCGCGCTGGCGCGG 10992

QY 130 CGAGCCGACGCGCGCGCGCGAGAGGCTGTGAGAGTGTGTGACAGAGACCCGCGGACAGA 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10993 ATGCTCGCGGCGCGCGCGCGGCGACTTCGACGCGGCGCTTCTCGGATCTGCGCGCGAGGCG 11052

QY 190 GGAACCATGGCTCCGCGAGAACTGTAGCA 217
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 11053 CTGGCGATGGAGCCCGCGAGCAGCGATCA 11080
```

Search completed: June 30, 2000, 13:32:40
Job time: 5921 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 11:55:33 ; Search time 70.1 Seconds

(Without alignments)
4911.046 Million cell updates/sec

Title: US-09-501-714-2

Perfect score: 1376
Sequence: 1 TCTACCGGAGCTCGGAGCT.....TTTATTTTCATATGCAACT 1376

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues

Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1374	99.9	1376	1 X06099	Human DnaJ-like pr
2	91.4	6.6	1330	1 X06100	Human DnaJ-like pr
3	70	5.1	1204	1 X14194	H. pylori GHP0 885
4	67.2	4.9	28295	1 X20507	Polynucleotide seq
5	65	4.7	672	1 T11788	Mouse cysteine str
6	65	4.7	672	1 T11789	Human cysteine str
7	62.6	4.5	1191	1 Q57429	DNA J like protein
8	60.6	4.4	110000	1 X20248_03	Continuation (4 of
9	59.4	4.3	110000	1 X20248_02	Continuation (3 of
10	59.2	4.3	1726	1 X07570	Homo sapiens fetal
11	57	4.1	1331	1 V86438	EST clone AM610. N
12	53.4	3.9	1515	1 Q77721	p58 CDNA clone. SC
13	53.4	3.9	1515	1 V22052	Bovine p58 DNA. SC
14	53.4	3.9	1687	1 V01072	Bovine p58 gene. S
15	49.4	3.6	6336	1 W74530	Staphylococcus aur
16	48	3.5	1700	1 V11858	Mus musculus Tubb 1
17	47.8	3.5	4320	1 T73388	DNA encoding HSP72
18	47.8	3.5	7174	1 V52322	Streptococcus pneu
19	47.8	3.5	110000	1 T58840_0	Mycoplasma genital
20	47.8	3.5	110000	1 T58840_2	Continuation (3 of
21	47.2	3.4	459	1 W87477	EST clone BN379. N
22	46.8	3.4	450	1 Q03633	Mycoplasma hyopneu
23	46.4	3.4	1701	1 X39655	Renal cancer assoc
24	46	3.3	828	1 X30633	H. pylori cytoplas
25	46	3.3	837	1 X30634	H. pylori cytoplas
26	46	3.3	990	1 X14165	H. pylori GHP0 542
27	45	3.3	2717	1 X13034	Enterococcus faeca
28	43.6	3.2	491	1 V02136	Human secreted pro
29	43.6	3.2	491	1 T88058	Partial CDNA clone
30	43.6	3.2	8169	1 V26609	Actinomodura hdbis
31	42	3.1	201	1 N70194	Signal portion of
32	41.8	3.0	336	1 X40691	Human secreted pro
33	41.6	3.0	201	1 N70195	Streptomyces prote
34	40.8	3.0	2312	1 X16152	Chicken Sox1 cDNA.

35	40.8	3.0	29879	1 Q46806	eryA region of S.
36	39.6	2.9	3748	1 T84330	Maize DnaJ-related
37	39.6	2.9	3748	1 V72905	Maize DnaJ clone D
38	39.2	2.8	832	1 X04354	DNA encoding a hum
39	38.8	2.8	1860	1 Q50632	Brain factor-2. Is
40	38.6	2.8	1055	1 T15275	E.coli K12 dnaJ co
41	38.6	2.8	5852	1 Q11710	Dictyostelium plas
42	38.4	2.8	284	1 V90426	EST clone DM360. N
43	38	2.8	2277	1 V05370	Human telomerase p
44	38	2.8	2277	1 V13834	Homo sapiens ambig
45	37.8	2.7	997	1 Q48024	ATGF gene #1. Cell

ALIGNMENTS

RESULT 1
X06099 standard; DNA; 1376 BP.
AC X06099: 196..1272
DT 09-APR-1999 (first entry)
DE Human DnaJ-like protein, HSPJ1 encoding DNA.
KW DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
KW immune disorder; inflammation; tissue damage; diabetes; wound healing;
KW chromosome mapping; ss.
OS Homo sapiens.
FH Key
FT Location/Qualifiers
FT 196..1272
FT /*tag= a
FT /product= "DnaJ-like protein, HSPJ1"
FT /transl_except= (pos:1087..1089; aa:Xaa)
FT /transl_except= (pos:1090..1092; aa:Xaa)
FT /note= "Xaa = unknown"

PD 10-DEC-1998.
PE 02-JUN-1998; U11182.
PR 03-JUN-1997; US-868288.
PA (INCY-) INCYTE PHARM INC.
PI Au-Young J, Bandman O, Lal P;
DR WPI: 99-070259/06.
DR P-PSDB: W94065.
PT New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
PT treatment and prevention of cancer, immune disorders and
PT inflammation
PS Claim 6; Fig 1A-D; 73pp; English.
CC This DNA encodes a human DnaJ-like protein, HSPJ1. The invention provides
CC two human DnaJ-like proteins which are heat shock proteins J1 and J2
CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
CC cells containing a vector comprising the nucleic acids are used for the
CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
CC to raise Ab, therapeutically and to screen for specific binding agents.
CC Antagonists are used to treat or prevent a wide variety of solid cancers,
CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC and parasitic infections) and inflammation. Agonists may be used to treat
CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC attack, ischemia, UV exposure, insulin-dependent diabetes and for wound
CC healing, and may be expressed from a vector containing the nucleic acids.
CC Fragments of the nucleic acids are used as primers and probes for
CC detecting and quantifying the HSPJ1-encoding nucleic acid in usual
CC hybridisation and/or amplification assays, therapeutically as antisense,
CC triplex-forming or ribozyme molecules, and for chromosome mapping.
SQ Sequence 1376 BP; 387 A; 268 C; 322 T;

Query Match 99.9%; Score 1374; DB 1; Length 1376;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1376; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCTACCGGAGCTCGGAGCTCCCGGAGTGCAGCGCAGAGAGGGGCTAGTACTG 60
DB 1 TCTACCGGAGCTCGGAGCTCCCGGAGTGCAGCGCAGAGAGGGGCTAGTACTG 60

[illegible][illegible]

OY	363	GAACCTCATGATGATCCACAGCCACAGAGAAATTCACAGATCTGGGGCTGCTTATAGGT	422
Db	140	AAACCCCGGCGCATTAAGAAGCGGAAGAAAATTCAACGCTCATCATGAAGCCTATGGGCT	199
OY	423	TCTGTCAGTACTGTGAAACCGGAAACGATGATCTTATGGTGAAGAAGATTAAAGA	482
Db	200	GTTAAGCGCATGAAAGAAAGCGGGCGCTTATACACAGGTATGTTAAAAAAGGCTTAAA	257
OY	483	TGCTCATCAGAGCTCCCATGGAGACATTTTTCACACTCTTGGGGGATTTTGGTTCA	542
Db	257	--CCAAACCGGCGCAAGCCAAAGGGGATTTTCTGATTTTGAAGATTATAGGCTGTT	313
OY	543	GTTTGGAG	550
Db	314	TTTTGAAG	321

RESULT 4

X20507

ID X20507 standard; DNA; 28295 BP.

AC X20507;

DT 05-MAY-1999 (first entry)

DE Polynucleotide sequence from the genome of *Treponema pallidum*.

KW *Treponema pallidum* infection; syphilis; *Borrelia* infection; animal;
enzyme production; ds.

OS *Treponema pallidum*.

PN W09859034-A2.

PD 30-DEC-1998.

PF 23-JUN-1998; U13041.

PR 24-JUN-1997; US-050667.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Fraser CM;

FI WPI: 99-081273/07.

PT New isolated *Treponema pallidum* nucleic acids - used to develop
products for the detection, diagnosis, characterisation, prevention
and therapy of *T. pallidum* infections, particularly syphilis

PS Claim 1; Page 210-227; 1150pp; English.

CC X20500-21243 represent polynucleotide sequences from the genome of
Treponema pallidum. The sequences can be used for detection,
diagnosis, characterisation, prevention and therapy for *T. pallidum*
infections, particularly syphilis. They can also be used for detecting
diseases related to *Borrelia* infections in animals, and for the
production of biosynthetic products such as enzymes.

Sequence 28295 BP; 6269 A; 6077 C; 8919 G; 6996 T;

Query Match	4.9%	Score 67.2	DB 1	Length 28295
Best Local Similarity	54.4%	Pred. No. 2.6e-08		
Matches 135	Conservative 0	Mismatches 113	Indels 0	Gaps 0
QY 241	CTCATCGGGGGCGGTCGATTGCCGCGACGAGATTCTCTAATGATCTTGGGGGTGCTTCGAAGT	300		
Db 26704	CTAGACGGGGTGATCGTGGCAAGAAAGGATTATTACGAGTTCTCGTATCTCAAGACC	26763		
QY 301	GCCCTATTAAGATATTTAAAAAGGCCCTATATAGAAACTAGCCCTGCAGCTTCATCCGAC	360		
Db 26764	GCGAGTGGAGAAAGTAATCAAAAAGGCGTACCGGCGGCTGCCTATTCAGTTTCATCTGAC	26823		
QY 361	CGGAACCCGTGATGATCCACAAAGCCACAGAGAAATTCAGAGATCTGGGTGCTGCTATGAG	420		
Db 26824	CGTATCTCAGGGAAATTAAGAGGCGGGAGGACGCTTCAAGAGGAGCTACCGAAGCCTATGAG	26883		
QY 421	GTTCTGTGAGATAGTAGAAGAAACGGAAGACGTACGATCTTATATGGAAGAAAGATTAAAA	480		
Db 26884	GTCGTCATTGATGACACAGAAAGCGTCCCGTCGATCGATGCGTTTGATGAGCCTGAAG	26943		
QY 481	GATGCTCA 488			
Db 26944	GATATGCA 26951			

	RESULT	6
ID	T11789	
AC	T11789	
DT	07-APR-1996	(first entry)
DE	Human cysteine string protein gene.	
KM	Human; cysteine string protein; anergy; T-lymphocyte;	
KW	differential display; cDNA; reverse transcription; probe;	
KW	polymerase chain reaction; cloning; gel electrophoresis;	
KW	antibody; diagnostic; immunoassay; autoimmune disease;	

RESULT	7
057429	
ID	057429 standard; cDNA to mRNA; 1191 BP.
AC	Q57428;
DT	19-OCT-1994 (first entry)
DE	DNA J like protein.
KW	Human cDNA; library; enzyme; protein; ss.
OS	Homo sapiens.
PN	WO9403599-A.
PD	17-FEB-1994.
PF	04-AUG-1993; J01095.
PR	04-AUG-1992; JP-208077.
PR	13-NOV-1992; JP-327619.
PR	26-FEB-1993; JP-061431.
PA	(SAGA) SAGAMI CHEM RES CENTRE.
PI	Iwahori A, Kato S, Kato T, Kim N, Oh S, Sekine S;
DR	WPI; 94-065688/08.
DT	P-PDB; R46090.
DT	cDNA of human origin and proteins coded by it - which may be

PT expressed by in vivo or in vitro translation using sense RNA or
 PT antisense DNA corresponding to the cDNA.

PS Claim 1; Page 44-47; 16pp; Japanese.

CC mRNA expressed in human fibrosarcoma cell line HT-1080 was

CC isolated and used to construct a cDNA library using vector

CC pKAI. Clone HP00067 encoding DNA J-like protein

CC was isolated.

CC Sequence 1191 BP; 397 A; 207 C; 309 G; 278 T;

Query Match 4.5%; Score 62.6; DB 1; Length 1191;
 Best Local Similarity 59.4%; Pred. No. 9.4e-08;

Matches 129; Conservative 0; Mismatches 79; Indels 9; Gaps 1;

QY 270 TTCTCTATAGATCTTGGGGGCTCGAAGTGCCTATATAAGATATTAAGAGCCCTA 329

Db 18 TTACTAGAGTCTTGGGGGCTCAAAACCAATGCTACAGAGAAATTAAGAGGCTTA 77

QY 330 TAGAACTAGCCCTGAGCTTCATCCGACCGAGACCTGATGATCCACAGCCCAAGA 389

Db 78 TAGAACTAGCCCTGAGCTTCATCCGAGACCTGATGATCCACAGCCCAAGA-----GA 128

QY 390 GAAATTCAGAGCTGCGTCTGTATGAGTCTGTCAAGATAGTGAAGAAAGCAACA 449

Db 129 GAAATTCAGAGCTGCGTCTGTATGAGTCTGTCAAGATAGTGAAGAAAGCAACA-----GA 188

QY 450 GTACGATCTTANGGTGAAGAAGATTAAGATGT 486

Db 189 ATATGACAAAGAGAGAGACAGCAATTAAGAGGT 225

RESULT 8

X20248_03

Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide s

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP X20248_00 1 110000

WP X20248_01 100001 210000

WP X20248_02 200001 310000

WP X20248_03 300001 410000

WP X20248_04 400001 510000

WP X20248_05 500001 610000

WP X20248_06 600001 710000

WP X20248_07 700001 810000

WP X20248_08 800001 910000

WP X20248_09 900001 910715

Query Match 4.4%; Score 60.6; DB 1; Length 110000;

Best Local Similarity 54.8%; Pred. No. 3.5e-06;

Matches 120; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 267 AGATTCTATAGATCTTGGGGGCTCGAAGTGCCTATATAAGATATTAAGAGCC 326

Db 81529 AGATTCTATAGATCTTGGGGGCTCGAAGTGCCTATATAAGATATTAAGAGCC 81588

QY 327 CTATAGAACTAGCCCTGAGCTTCATCCGACCGAGACCTGATGATCCACAGCCCA 386

Db 81589 TTATAGAACTAGCCCTGAGCTTCATCCGACCGAGACCTGATGATCCACAGCCCA 81648

QY 387 GGAGAAATTCAGAGTCTGGTCTGTATGAGTCTGTCAAGATAGTGAAGAAAGCA 446

Db 81649 CTATATCTTTAAAGAACCACTCAGCTTACGAAATTTATATGATGACATATAAAGGC 81708

QY 447 ACAGTACGATCTTATGATGTAAGAAGATTAAGATG 485

Db 81709 TAAATAGCAGATTTGGGCACTTCGCTTTGAAGAGG 81747

RESULT 9

X20248_02

Continuation (3 of 10) of X20248 from base 200001 (Borrelia burgdorferi polynucleotide s

WP Sequence split into 10 fragments LOCUS X20248 Accession X20248

WP X20248_00 1 110000

WP X20248_01 100001 210000

WP X20248_02 200001 310000

WP X20248_03 300001 410000

WP X20248_04 400001 510000

WP X20248_05 500001 610000

WP X20248_06 600001 710000

WP X20248_07 700001 810000

WP X20248_08 800001 910000

WP X20248_09 900001 910715

Query Match 4.3%; Score 59.4; DB 1; Length 110000;

Best Local Similarity 51.3%; Pred. No. 7.5e-06;

Matches 138; Conservative 0; Mismatches 131; Indels 0; Gaps 0;

QY 267 AGATTCTATAGATCTTGGGGGCTCGAAGTGCCTATATAAGATATTAAGAGCC 326

Db 15206 AGACTACTACATATCTTGGGATACAAAAAATGCTATGATGAGAAATTAAGAGC 15265

QY 327 CTATAGAACTAGCCCTGAGCTTCATCCGACCGAGACCTGATGATCCACAGCCCA 386

Db 15266 TTACAAAAAATGGCAATTAATATCCAGACGACAAAAACAGGAAACAAATAGCTGA 15325

QY 387 GGAAGAAATTCAGAGTCTGGTCTGTATGAGTCTGTCAAGATAGTGAAGAAAGCA 446

Db 15326 AGAAAGATTTAAAGAAATTAATGAGCTTATGAATTTATCTCTCTCTATTAAGAAAG 15385

QY 447 ACAGTACGATCTTATGATGTAAGAAGATTAAGATGATGATGATGATGATGATG 506

Db 15386 AATATATGACTCTTGGGATACAAAAATTTAAAGGCAACAGCAATTTTGAAGAGGA 15445

QY 507 CATTTTTCACACTCTTGGGATTTTG 535

Db 15446 ATTTAGCAGCAGCAAGATTTGGCAATTTTG 15474

RESULT 10

X07570

ID X07570 standard; cDNA; 1726 BP.

AC X07570:

DT 21-JUN-1999 (first entry)

DE Homo sapiens fetal kidney clone AM610 secreted protein gene.

KW Secreted protein; fetal kidney; ds.

OS Homo sapiens.

FH Key Location/Qualifiers

FT CDS 30..701

FT /*tag= a

FT /product= "secreted protein"

PN W09900405-A1.

PD 07-JAN-1999.

PR 29-JUN-1998; UI3530.

PR 30-JUN-1997; US-885610.

PA (GEMT) GENETICS INST INC.

PI Agostino MJ, Evans C, Jacobs K, Lavallie ER, Mccoy JM,

DR P-PSDB: W95710.

DR P-PSDB: W95710.

PT New polynucleotides encoding secreted human proteins - are derived

PT from foetal kidney or adult retina cDNA libraries, used as, e.g.

PT potential vaccines

PS Disclosure: Pages 55-56; 76pp; English.

CC The sequence is that encoding a secreted protein from a human fetal

CC kidney clone AM610. Such a sequence is predicted to have biological

CC activities which would make them suitable for treating, preventing or

CC ameliorating medical conditions in humans and animals, although no

CC supporting data is given. Suggested activities include nutritional

CC activity, cytokine and cell proliferation/differentiation activity,

CC immune stimulating (e.g. as vaccines) or suppressing activity,

CC haematopoiesis regulating activity, tissue growth activity,

CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic

CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory

CC activity, cadherin/tumour invasion suppressor activity, and tumour
CC inhibition activity. It is also stated to be useful for gene
CC therapy.
S0 Sequence 1726 BP; 543 A; 285 C; 315 G; 583 T;

Query Match 4.3%; Score 59.2; DB 1; Length 1726;
Best Local Similarity 51.2%; Pred. No. 9.9e-07;
Matches 164; Conservative 0; Mismatches 153; Indels 3; Gaps 1;

QY 169 TGAAGACAGACCCGAGAGAAACCATGGCTCCGAGAACGAGACCTTTGGCTG 228
DB 6 TGGACAGAGGCGAGATATTAGAAATGGCTACTCCCGCATCAATTTTCATCTTGCATC 65
QY 229 TTGCTGTATACCTCATCGGGCGGTGATGCGGAGAGATTTCTATTAAGATCTGGGG 288
DB 66 TGCATTTTATGATTAACAGATTAATCTGCGCTCAAAAACCTACTATGATATCTTAGGT 125
QY 289 GTGGCTGGAAGTGCCTCTATTAAGATTTTAAAAAGCCCTATAGAAAAGCTAGCCCTGCAG 348
DB 126 GTGCAAAATCGGATCAGACCGCCAAATCAAGAGGCTTTCAAGTTGGCATGAAG 185
QY 349 CTTTATCCCGACGAAACCTGATGATCAACAGCCAGAGAAATTCAGAGATCTGGGT 408
DB 186 TACACCCCTGACAAAA---TAAAGCCAGATCTGAGCAAAATTCAGAGATTCGA 242
QY 409 GCTGCTTATGAGTGTCTGATGATAGTGAAGAAAGAGTACGATCTTATGCTGAA 468
DB 243 GAAGCATATGAAACATCTCAGATGTAATAGCAAAAGATGATACACTTGACAC 302
QY 469 GAAGGATTAAGATGTGCA 488
DB 303 AGTCTTTTACTAGTGTAA 322

RESULT 11

VS6438
ID VS6438 standard: cDNA: 331 BP.
AC VS6438;
DE 27-APR-1999 (first entry)
KW Expressed sequence tag: secreted protein; haematopoiesis regulator;
KW tissue growth; activin; inhibin; tumour invasion suppressor; EST; human;
KW chemokinesis; chemokinesis; haemostasis; gene therapy; thrombolytic;
KW receptor; ligand; anti-inflammatory; tumour inhibitor; ds.
OS Homo sapiens.
PN W09845435-A2.
PD 15-OCT-1998.
PR 10-APR-1998; U06954.
PR 10-APR-1997; US-835913.
PA (GENE) GENETICS INST INC.
PI Agostino M, Jacobs K, Lavallie ER, McCoy JM, Merberg D,
PI Ragle LM, Spaulding V, Treacy M,
PI WPI: 99-070076/06.
DR New polynucleotides encoding human secreted proteins - derived from
PT e.g. human blood, kidney, foetal lung, placenta, testes, brain,
PT ovary, pituitary, retina and colon cDNA libraries
PS Claim 1, Page 234, 633pp; English.
CC This sequence represents an expressed sequence tag (EST), and is a
CC polynucleotide of the invention. The polynucleotides of the invention are
CC all secreted EST sequences isolated from a variety of human tissue
CC sources. The EST sequences and proteins encoded by them are predicted to
CC have useful biological activities which would make them suitable for
CC treating, preventing or ameliorating medical conditions in humans and
CC animals, although no supporting data is given. Suggested activities
CC include nutritional activity, immune stimulating or suppressing activity,
CC haematopoiesis regulating activity, tissue growth activity,
CC activin/inhibin activity, chemokinesis/chemokinesis activity, haemostatic
CC activity, cadherin/tumour invasion suppressor activity, tumour inhibition
CC activity. The EST sequences are also stated to be useful for gene
CC therapy.
S0 Sequence 331 BP; 115 A; 69 C; 68 G; 79 T;

Query Match 4.1%; Score 57; DB 1; Length 331;
Best Local Similarity 52.2%; Pred. No. 1.7e-06;
Matches 151; Conservative 0; Mismatches 135; Indels 3; Gaps 1;

QY 200 CTCGAGAACCTGAGACCTTTGGCTGTGCGCTATACCTCATGCGGGCGGTATTTG 259
DB 36 CTCGCCAGTCAATTTTCATCTTGCATTCGATTTTAAATGATTAAGATTAATCTCG 95
QY 260 CCGAGCAGATTTCTATTAAGATCTTGGGGGTGCTCGAATGCTCTATTAAGATTA 319
DB 96 CTCGAAAGCTACTATGATATATCTAGTGTGCAAAATGCGATCAGACGCCAAATCA 155
QY 320 AAAAGCCCTTATAGAACTAGCCCTGAGCTTATATCCAGCCGAACCTGATGATCCAC 379
DB 156 AGAAGGCTTTACCAAACTGCGCATGAAGTACACACCTGACAAAA---TAAAGCCCA 212
QY 380 AACCCAGAGAAATTCAGATCTGGGTGCTGCTTATAGAGTGTCTGAGATAGAGA 439
DB 213 ATGCTGAGCAAAATTCAGAGATTCGAGAAACATATGAAACATCTCAGATGTAATA 272
QY 440 AACGAAACAGTACGATCTTATGATGTAAGAAAGATTAAGATGTCA 488
DB 273 GACGAAAGATGATGATACACTTGACACAGTCTTTTACTAGTGTAA 321

RESULT 12

Q77721
ID Q77721 standard: cDNA: 1515 BP.
AC Q77721;
DE 05-JUL-1995 (first entry)
DE p58 cDNA clone.
KW Influenza virus; p58; probe; MDBK cell line; RACE-PCR; ss.
OS Influenza virus.
PN W09423041-A.
PD 13-OCT-1994.
PR 01-APR-1994; U03623.
PR 02-APR-1993; US-042024.
PA (RIBO-) RIBOGENE INC.
PI Katze MG, Mathews MB, Miles VJ, Watson JC, Withereil G;
PI WPI: 94-333201/41.
DR Screening for antiviral agents which have little effect on
PT non-infected cells - by determining whether the agent interacts
PT with a virus or cellular component and allows or prevents
PT preferential translation of viral RNA.
PS Example 6: Page 168-69; 195pp; English.
CC This sequence represents the Influenza virus p58 cDNA. This sequence
CC was isolated using probes based on tryptic peptides derived from p58.
CC These probes were used to screen a cDNA library from MDBK cell line
CC made in lambda Zap II vector. The initial clone contained a long
CC open reading frame but no termination codon suggesting that the 3'
CC end was missing. The 3' end was isolated by RACE-PCR. This allowed
CC construction of the full length p58 cDNA containing 1680 bp.
S0 Sequence 1515 BP; 509 A; 285 C; 362 G; 359 T;

Query Match 3.9%; Score 53.4; DB 1; Length 1515;
Best Local Similarity 58.1%; Pred. No. 3.7e-05;
Matches 118; Conservative 0; Mismatches 76; Indels 9; Gaps 1;

QY 263 GACGAGATTTCTATTAAGATCTTGGGGGTGCTCGAAGTGCCTTATTAAGATTAATA 322
DB 1175 GACGAGATTTATTAACAATCTTGGAGATTAAGAAATGSCAAAGCAAGCAATATTA 1234
QY 323 AGGCTATAGAAATATACCCCTGCGAGCTTCATCCGACCGAACCCTGATATCCAC--- 380
DB 1235 AAGCATACCGAAATTAACCTGAGTGACACCAAGAACCTTCAGAAAGAAAGAA 1294
QY 380 -----AAGCCAGAGAAATTCAGATCTGGGGGTGCTTATAGAGTGTGTCAGATA 433
DB 1295 AGAAAGAGCTGAGAAAGATTCATGACATAGACAGCTGCTTAAGAAAGTCTTCGATC 1354

given in the specification for this DNA sequence"

FT given in the specification for this DNA sequence
 PN EP-786519-AZ.
 PD 30-JUL-1997.
 PF 07-JAN-1997. 100117.
 PR 05-JAN-1996; US-009861.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA,
 PI Rosen CA;
 PI WPI: 97-374922/35.
 PR Polynucleotide(s) and proteins derived from *Staphylococcus aureus* -
 PR stored on computer readable medium and used in the production of
 PR anti-S.aureus vaccines
 PS Claim 1, Page 969-972, 3271pp; English.
 CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
 CC of the invention. The DNA sequences are recorded on a computer readable
 CC medium, preferably selected from a floppy or hard disk, random access
 CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
 CC the S.aureus DNA sequences allows putative homologs to be assigned so
 CC that protein-encoding or regulatory regions of commercial, therapeutic or
 CC industrial importance can be obtained. Specifically, sequences which are
 CC likely to encode antigens have been identified and these polypeptides can
 CC be used in a vaccine composition against S.aureus infection. The
 CC polypeptides can also be used in a kit for the immunodetection of
 CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,
 CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
 CC skin and surgical wound infections, scalded skin syndrome, toxic shock
 CC syndrome, etc. Organisms transformed with the DNA sequences can be used
 CC for recombinant production of the polypeptides. The new DNA sequences
 CC (and their fragments) are useful as primers or probes for isolating
 CC homologues of any of the S.aureus DNA sequences contained on the
 CC computer readable medium.
 SQ Sequence 6336 BP: 1733 A; 1218 C; 839 G; 2364 T;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:57:47 ; Search time 16.52 Seconds

(without alignments)
288.398 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGLGVQRHNSPEDIK.....EAEKGVEEEDVQROSLDRT 330

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 145308 seqs, 14437401 residues

Total number of hits satisfying chosen parameters: 145308

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696	98.8	330	2	US-08-868-288A-3
2	1696	98.8	330	3	US-09-225-373-3
3	1696	98.8	330	3	US-09-388-993-3
4	635.5	37.0	277	2	US-08-868-288A-7
5	635.5	37.0	277	3	US-09-235-373-7
6	635.5	37.0	277	3	US-09-388-993-7
7	635.5	37.0	277	3	US-09-388-993-7
8	635.5	37.0	351	2	US-08-868-288A-6
9	635.5	37.0	351	3	US-09-235-373-6
10	364.5	21.2	340	2	US-08-974-546-5
11	337	19.6	348	2	US-08-974-546-1
12	311	18.1	397	2	US-08-868-288A-5
13	311	18.1	397	3	US-09-235-373-5
14	311	18.1	397	3	US-09-388-993-5
15	274.5	16.0	419	2	US-08-686-417-3
16	259.5	15.1	352	2	US-08-472-534-6
17	257	15.0	358	2	US-08-868-288A-1
18	257	15.0	358	3	US-09-235-373-1
19	257	15.0	358	3	US-09-388-993-1
20	241.5	14.1	484	3	US-08-879-260-4
21	232.5	13.5	438	2	US-08-897-340-34
22	178	10.4	51	1	US-08-346-849-12
23	178	10.4	51	1	US-08-293-284A-12
24	171.5	10.0	52	1	US-08-346-849-10
25	171.5	10.0	52	2	US-08-293-284A-10
26	160.5	9.4	52	1	US-08-346-849-8
27	160.5	9.4	52	2	US-08-293-284A-8
28	151.5	8.8	52	1	US-08-346-849-9
29	151.5	8.8	52	2	US-08-293-284A-9

30	147	8.6	49	1	US-08-346-849-11	Sequence 11, Appl
31	147	8.6	49	2	US-08-293-284A-11	Sequence 11, Appl
32	110.5	6.4	56	2	US-08-346-849-28	Sequence 28, Appl
33	110.5	6.4	56	2	US-08-293-284A-28	Sequence 28, Appl
34	105	6.1	51	1	US-08-346-849-14	Sequence 14, Appl
35	105	6.1	51	2	US-08-293-284A-14	Sequence 14, Appl
36	103.5	6.0	341	2	US-08-538-711A-8	Sequence 8, Appl
37	103.5	6.0	353	2	US-08-538-711A-7	Sequence 7, Appl
38	103.5	6.0	433	1	US-08-346-849-2	Sequence 2, Appl
39	103.5	6.0	433	2	US-08-293-284A-2	Sequence 2, Appl
40	103.5	6.0	714	2	US-08-990-114-3	Sequence 3, Appl
41	98	5.7	223	1	US-07-667-276A-10	Sequence 10, Appl
42	96.5	5.6	643	2	US-08-551-356-4	Sequence 4, Appl
43	96.5	5.6	643	4	PCT-US93-12687-4	Sequence 4, Appl
44	96	5.6	51	1	US-08-346-849-15	Sequence 15, Appl
45	96	5.6	51	2	US-08-293-284A-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-868-288A-3
: Sequence 3, Application US/08868288A
: Patent No. 5922567
: GENERAL INFORMATION:
: APPLICANT: Au-Young, Janice
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESS: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/868,288A
: FILING DATE: June 3, 1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0309 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 330 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: HNT2RAT01
: CLONE: 260873
: US-08-868-288A-3

Query Match 98.8%; Score 1696; DB 2; Length 330;
Best Local Similarity 100.0%; Pred. No. 2.9e-167;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDYEVGLGVQRHNSPEDIKKAYRKALKWHDKPKKEAEKRFQYAEYEVLSDAK 60

```
|||||
Db 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEAYEVLSDAK 60
QY 61 KRDIYDKGKGLNGGGGSHFDSPEFEGFTFRNPDVDFEFGRGRPFSEDFEDPE 120
Db 61 KRDIYDKGKGLNGGGGSHFDSPEFEGFTFRNPDVDFEFGRGRPFSEDFEDPE 120
QY 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
Db 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
QY 181 GSGGNGFKSISTSTKMNNGRKITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
Db 181 GSGGNGFKSISTSTKMNNGRKITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
QY 241 ERMRRGQNVLPAPAGLRPPKPPRPSLLRHXPHCLSKSEEGEDRPMAPXXMPLASXAG 300
Db 241 ERMRRGQNVLPAPAGLRPPKPPRPSLLRHXPHCLSKSEEGEDRPMAPXXMPLASXAG 300
QY 301 XXEGKKRMAEAERGVEEVDQROSLDRT 330
Db 301 XXEGKKRMAEAERGVEEVDQROSLDRT 330
```

```
RESULT 2
US-09-235-373-3
; Sequence 3, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/668,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-235-373-3
```

Query Match 98.8%; Score 1696; DB 3; Length 330;
Best Local Similarity 100.0%; Pred. No. 2,9e-167;

```
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEAYEVLSDAK 60
Db 1 MVDYEVLVGVRHASPEDIKAYRKIALKWHDPKNPENKEAEARKKQVAEAYEVLSDAK 60
QY 61 KRDIYDKGKGLNGGGGSHFDSPEFEGFTFRNPDVDFEFGRGRPFSEDFEDPE 120
Db 61 KRDIYDKGKGLNGGGGSHFDSPEFEGFTFRNPDVDFEFGRGRPFSEDFEDPE 120
QY 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
Db 121 DFGNRRGPRGSRSGTSGFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSGF 180
QY 181 GSGGNGFKSISTSTKMNNGRKITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
Db 181 GSGGNGFKSISTSTKMNNGRKITTKRIYENGQERVEVEDGQLSLTINGVADDALXE 240
QY 241 ERMRRGQNVLPAPAGLRPPKPPRPSLLRHXPHCLSKSEEGEDRPMAPXXMPLASXAG 300
Db 241 ERMRRGQNVLPAPAGLRPPKPPRPSLLRHXPHCLSKSEEGEDRPMAPXXMPLASXAG 300
QY 301 XXEGKKRMAEAERGVEEVDQROSLDRT 330
Db 301 XXEGKKRMAEAERGVEEVDQROSLDRT 330
```

```
RESULT 3
US-09-388-993-3
; Sequence 3, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/668,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 330 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: HNT2RAT01
; CLONE: 260873
US-09-388-993-3
```

Query Match 98.8%; Score 1696; DB 3; Length 330;
 Best Local Similarity 100.0%; Pred. No. 2.9e-167;
 Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDNKPNKEEAEKRFQVAAYEVLSDAK 60
 |||||
 Db 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDNKPNKEEAEKRFQVAAYEVLSDAK 60
 QY 61 KRDIYDKYKGBLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 120
 |||||
 Db 61 KRDIYDKYKGBLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 120
 QY 121 DFGNRRGPGSRGSRGSGFSAFSGPSSGSSFDGTGTSGLGHGGLTSFSSSF 180
 |||||
 Db 121 DFGNRRGPGSRGSRGSGFSAFSGPSSGSSFDGTGTSGLGHGGLTSFSSSF 180
 QY 181 GGSAGMFKSISTSTKMYNGRKITTKRIVENGEVEVEEDGQLKSLTINGVADDDALXE 240
 |||||
 Db 181 GGSAGMFKSISTSTKMYNGRKITTKRIVENGEVEVEEDGQLKSLTINGVADDDALXE 240
 QY 241 ERMRRGQNVLPAPAGLRPPAPASLLRHXPCHLSKEEGDQRPMAPIXWPLASXAG 300
 |||||
 Db 241 ERMRRGQNVLPAPAGLRPPAPASLLRHXPCHLSKEEGDQRPMAPIXWPLASXAG 300
 QY 301 XXEGXKRMXAEAEKGVVEEVDQROSLDRT 330
 |||||
 Db 301 XXEGXKRMXAEAEKGVVEEVDQROSLDRT 330

RESULT 4

US-08-868-288A-7
 ; Sequence 7, Application US/08868288A
 ; Patent No. 5922567

GENERAL INFORMATION:
 APPLICANT: Au-Young, Janice
 APPLICANT: Lal, Preeti
 TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/868,288A
 FILING DATE: June 3, 1997
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 277 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank

CLONE: 32470
 US-08-868-288A-7

Query Match 37.0%; Score 635.5; DB 2; Length 277;
 Best Local Similarity 54.4%; Pred. No. 7.2e-58;
 Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALKWHPDNKPNKEEAEKRFQVAAYEVLSDAK 60
 |||||
 Db 1 MASYEILIDVRSASADIKAYRKALKWHPDNKPNKEEAEKRFQVAAYEVLSDAK 60
 QY 61 KRDIYDKYKGBLNGGGGGSHEDSPFEFGTFRNPDVREFFGADPFSEDFEEDPFE 117
 |||||
 Db 61 KREIYDRGRGLTGTGTPRAEAGSGPGFTTFNPSPEEVEVEEDGQDPFA-ELFDD 119
 QY 118 --PEDEFGNRRGPGSRGSRGSGFSAFSGPSSGSSFDGTGTSGLGHGGLTSF 175
 |||||
 Db 120 LGPSEL--QNRG-----SRHSGPFTTSSSP-----GH--SDF 150
 QY 176 SSTSFGGS--GKGNFKSISTSTKMYNGRKITTKRIVENGEVEVEEDGQLKSLTINGVAD 234
 |||||
 Db 151 SSSFSFGAGATRSYSTTFVQGRITTRIMENGEVEVEEDGQLKSLTINGVAD 210
 QY 235 DDALXEERMRGQNVLPAPQA 255
 |||||
 Db 211 DLANGELSRREQ-----QPS 226

RESULT 5

US-09-235-373-7
 ; Sequence 7, Application US/09235373
 ; Patent No. 6001598

GENERAL INFORMATION:
 APPLICANT: Au-Young, Janice
 APPLICANT: Lal, Preeti
 TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
 NUMBER OF SEQUENCES: 7
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Drive
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/235,373
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/868,288
 FILING DATE: June 3, 1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0309 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 415-855-0555
 TELEFAX: 415-845-4166
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 277 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: GenBank
 CLONE: 32470

US-09-235-373-7

Query Match	37.08;	Score 635.5;	DB 3;	Length 277;
Best Local Similarity	54.48;	Pred. No. 7.2e-58;		
Matches 142;	Conservative 30;	Mismatches 48;	Indels 41;	Gaps 9

```

QY 1 MVDYEVGVORVHASPEDIKKAYKRLKMHDPDKNPNNKEAEKKKQVAAEYVUSDAK 60
    | | | | | : | : | | | | | : | | | | | : | | | | | : | | | | |
Db 1 MASYEILIDVRSASADDIKKAYIRKALOMHPDKNPDKKEAEKKKEVAAEAEVUSDKH 60
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 61 KRDIYDKYKGEGLNGGGGGCGSHDPSF--EFGFTFRNPDPVFEKEEFGGADPESFDFFED 117
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 61 KREIYDRGREGGLGTGTGSPRAAAGGGCPQFTTFPSPEEVEFRFGSDPFA-ELFDD 119
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 118 --PEDFDGNGRGGRGSRNRGTSGFFSAFSGFPBFGSGFSFDGTGTFSGSLGHGCLTSD 175
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 120 LGPFSSEL--QNRG----SRHSGPFFTFSSSP-----GH--SDF 150
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 176 SSTSGFSGS-GGNGFNKSISTSTKMWNGRRITTKRIVENGQERVEEEDGOLKSLTINGVAD 234
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 151 SSSFSFSPGAGARSVSTSTTFVQGRITTRRMENGGQERVEEEDGOLKSLTINGVPD 210
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
QY 235 DDALXERMRRGQNVLPAGQA 255
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |
Db 211 DLARGLELSRREQ-----QPS 226
    | | | | | : | | | | | : | | | | | : | | | | | : | | | | |

```

RESULT 6

```

US-09-388-993-7
:
: Sequence 7, Application US/09388993
: Patent No. 6043222
: GENERAL INFORMATION:
: APPLICANT: Au-Young, Janice
: APPLICANT: Lal, Preeti
: APPLICANT: Bandman, Olga
: TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
: NUMBER OF SEQUENCES: 7
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Incyte Pharmaceuticals, Inc.
: STREET: 3174 Porter Drive
: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/388,993
: FILING DATE:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/868,288
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Billings, Lucy J.
: REGISTRATION NUMBER: 36,749
: REFERENCE/DOCKET NUMBER: PF-0309 US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 415-855-0555
: TELEFAX: 415-845-4166
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 277 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: LIBRARY: GenBank
: CLONE: 32470
:
: US-09-388-993-7

```

Query Match	37.0%;	Score 635.5;	DB 3;	Length 277;
Best Local Similarity	54.4%;	Pred. No. 7.2e-58;		
Matches 142;	Conservative 30;	Mismatches 48;	Indels 41;	Gaps 9;

[illegible]

RESULT 7
HE-08-060

```

US-08-868-288A-6
? Sequence 6, Application US/08868288A
? Patent No. 5922567
? GENERAL INFORMATION:
? APPLICANT: Au-Young, Janice
? APPLICANT: Lal, Preeti
? APPLICANT: Bandman, Olga
? TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE
? NUMBER OF SEQUENCES: 7
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Incyte Pharmaceuticals, Inc.
? STREET: 3174 Porter Drive
? CITY: Palo Alto
? STATE: CA
? COUNTRY: USA
? ZIP: 94304
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Diskette
? COMPUTER: IBM Compatible
? OPERATING SYSTEM: DOS
? SOFTWARE: FastSeq for Windows Version 2.0
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/868,288A
? FILING DATE: June 3, 1997
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME: Billings, Lucy J.,/49
? REGISTRATION NUMBER: 36,749
? REFERENCE/DOCKET NUMBER: PF-0309 US
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 415-855-0555
? TELEFAX: 415-845-4166
? INFORMATION FOR SEQ. ID NO: 6:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 351 amino acids
? TYPE: amino acid
? STRANDEDNESS: single
? TOPOLOGY: linear
? IMMEDIATE SOURCE:
? LIBRARY: GenBank
? CLONE: 32469
US-08-868-288A-6

```


Best Local Similarity 54.4%; Pred. No. 1e-57;
Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVLGVRHASPEDIKKAYRKALWKHPDKNPENKEEAEKFKOVAAEAYVLSDAK 60
Db 1 MASTYELDPKRSADIKKAYRKALWKHPDKNPENKEEAEKFKOVAAEAYVLSDAK 60
QY 61 KRDIYDKYEGEGINGGGGSHFDSPP---EFQFTFRNPDPVFEFGGHDPEFSDFED 117
Db 61 KREIYDYRGEGGLGTGTGSRRAEAGSGGFTTFRSPPEVFRFFGSDPFA-ELFDD 119
QY 118 --FPEDFGRRGPRGSRSGTGSFSAFSGFSGFSFDTGTFSGSLGHGILTSP 175
Db 120 LGPFSEL--QNRG-----SRHSGPFFTFSSSF-----GH---SDF 150
QY 176 SSTFSGGS--GMGNFKSISTSTKMWNGRKITTKRIYENGQERVEVEEDGOLKSLTINGVAD 234
Db 151 SSSFSFSPAGAFRSYSTTTFVQGRITTRIMENGQERVEVEEDGOLKSLTINGVAD 210
QY 235 DDALXERMRGQNVLPAPPA 255
Db 211 DLARGELSRREQ-----QPS 226

RESULT 10
US-08-974-546-5

; Sequence 5, Application US/08974546
; Patent No. 5945287

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/974,546

; FILING DATE: Filed Herewith

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0428

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 340 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: GenBank

; CLONE: 1816452

Best Local Similarity 32.9%; Pred. No. 9.2e-30;
Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DYIEVLGVORHASPEDIKKAYRKALWKHPDKNPENKEEAEKFKOVAAEAYVLSDAK 62
Db 4 DYIOTGLAGADDEIKKAYRKALWKHPDKNPENKEEAEKFKOVAAEAYVLSDAK 61
QY 63 DIYDKYEGEGINGGGGSHFDSPPFEFGTFR-NPDVFEFGGHDPEFSDFED 116
Db 62 EIDRKGEEGLKSGSGSGGANGTS---SYTHGDHAFAEFGGR----- 110
QY 117 DPEDFGRRGPRGSRSGTGSFSAFSGFSGFSFDTGTFSGSLGHGILTSP 176
Db 110 NPEDTFEGQNRGEGM-----DIDDFSGFP-----MGMG---FT 142
QY 177 STFGSGSGMGNFKSISTSTKMWNGRKITTKRIYENGQERVEVEE--DGOLKSLTINGVAD 231
Db 143 NVNFGRS-----RSAQEPARKKODPPVT-----HDLVSLSEIYSGCTKKIKISHRKL 190
QY 231 -----GVADDALXERMRG-----QNVLPAPPAQRLRPKPPRPAASLRHXPCHSKRE 280
Db 191 NPDKSIRNEDKILTEVKGKREGKITTPKGGDOTSNMIPADIVLKDKNHLEKRD 250
QY 281 G 281
Db 251 G 251

RESULT 11

US-08-974-546-1

; Sequence 1, Application US/08974546

; Patent No. 5945287

; GENERAL INFORMATION:

; APPLICANT: Hillman, Jennifer L.

; APPLICANT: Lal, Preeti

; APPLICANT: Corley, Neil C.

; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Incyte Pharmaceuticals, Inc.

; STREET: 3174 Porter Dr.

; CITY: Palo Alto

; STATE: CA

; COUNTRY: USA

; ZIP: 94304

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FASTSEQ for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/974,546

; FILING DATE: Filed Herewith

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Billings, Lucy J.

; REGISTRATION NUMBER: 36,749

; REFERENCE/DOCKET NUMBER: PF-0428

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 650-855-0555

; TELEFAX: 650-845-4166

; INFORMATION FOR SEQ ID NO: 1:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 348 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; IMMEDIATE SOURCE:

; LIBRARY: BRATUT721

; CLONE: 2525691

Query Match 21.2%; Score 364.5; DB 2; Length 340;

US-08-974-546-1

Query Match 19.6%; Score 337; DB 2; Length 348;
Best Local Similarity 32.2%; Pred. No. 6.6e-27;
Matches 96; Conservative 39; Mismatches 101; Indels 62; Gaps 11;

QY 3 DYVYLVQVRASPEDIKKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 62
| | | | | : : : | | | | | : | | | | | : | | | | | : | | | | | :
DB 4 DYVYLVQVRASPEDIKKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 61
QY 63 DYVYLVQVRASPEDIKKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 119
| | | | | : : : | | | | | : | | | | | : | | | | | : | | | | | :
DB 62 GLVYQVEEGIKTGSGSG-----SSGSHFTTFHGDPHATASFFG-----SNPF 108
QY 120 EDFGNRRGPGSRSGTGSFSAFSGFSGFSGFSGFSGFSGFSGFSGFSGFSG 179
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 109 DIFGASRSTR-----PFGSPDDMDVDEDEDFGAFGRGFGNLSRGPARRA 156
QY 180 FGGSGMGNFSGTSGTSMVNGRKITTKRIYNGQRYVEE--DGQLKSLTI----- 230
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 157 -----PEPLYPARKVQDDPPVH--ELRVSLERIVHSGTRMKITRRRLND 200
QY 230 -NGVADDALXERRRG-----QNVLPAPAGLRPPRPPASILRRXPHCLSKKEG 281
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 201 GRVYTEDKILHIVYIKGWMKEGKITTPPKSGDAPDIPADIVFLKDKPHAFRRRDG 258

RESULT 12

US-08-868-288A-5
; Sequence 5, Application US/08868288A
; Patent No. 5922567
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/868,288A
; FILING DATE: June 3, 1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 306714
; US-08-868-288A-5

Query Match 18.1%; Score 311; DB 2; Length 397;
Best Local Similarity 32.8%; Pred. No. 3.8e-24;
Matches 88; Conservative 30; Mismatches 72; Indels 78; Gaps 10;

QY 4 YVEVYGVORHASPEDIKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 63
| | | | | : : : | | | | | : | | | | | : | | | | | : | | | | | :
DB 7 DYVYLVQVRASPEDIKKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 62
QY 64 YVYVYGVORHASPEDIKAYRKALKMHPKPNENKEAEARKQVAAEVLSDAKR 115
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 63 LYDKGGEARINEGGAGG-----FSSPDIFPMFTGGGGRMQRERKGNVYHQL 111
QY 116 EDPEDEFFG-----NRGPRGS-----RSRGTGSAFSGFSGFSG 152
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 112 SVTLEDLYNGATRKALKQKNVICDKCEGRGKKGAVGCCPNCRGQM1----- 161
QY 153 GFSSPDFTGTFSGLSGHGLTFSSTSGSGMGNFSGTSMVNGRKITTKRIYEN 211
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 161 -----RIHQIGPMVOQIOSVCMCEGQGHERRISPKDRCKSCNGKITVREKILE- 210
QY 212 GOERYEVE---EDGQLKSLTINGVADD 236
| | | | | : | | | | | : | | | | | : | | | | | : | | | | | :
DB 210 ----VHIDKMKDQ--KITFHGEGDQ 231

RESULT 13

US-09-235-373-5
; Sequence 5, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 397 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 306714
; US-09-235-373-5

? PRIOR APPLICATION DATA:
 ? APPLICATION NUMBER: 08/868,288
 ? FILING DATE:
 ? ATTORNEY/AGENT INFORMATION:
 ? NAME: Billings, Lucy J.
 ? REGISTRATION NUMBER: 36,749
 ? REFERENCE/DOCKET NUMBER: PF-0309 USA
 ? TELECOMMUNICATION INFORMATION:
 ? TELEPHONE: 415-855-0555
 ? TELEFAX: 415-845-4166
 ? INFORMATION FOR SEQ ID NO: 5:
 ? SEQUENCE CHARACTERISTICS:
 ? LENGTH: 397 amino acids
 ? TYPE: amino acid
 ? STRANDEDNESS: single
 ? TOPOLOGY: linear
 ? IMMEDIATE SOURCE:
 ? LIBRARY: GenBank
 ? CLONE: 306714
 ? US-09-388-993-5

```

1      FILING DATE: 26-JUL-1996
2      CLASSIFICATION: 800
3      ATTORNEY/AGENT INFORMATION:
4      NAME: Mursahlgie, Kate H.
5      REGISTRATION NUMBER: 29, 959
6      REFERENCE/DOCKET NUMBER: 27112-20038.000
7      TELECOMMUNICATION INFORMATION:
8      TELEPHONE: (202) 887-1500
9      TELEFAX: (202) 822-0168
10     INFORMATION FOR SEQ ID NO: 3:
11     SEQUENCE CHARACTERISTICS:
12     LENGTH: 419 amino acids
13     TYPE: amino acid
14     TOPOLOGY: linear
15     MOLECULE TYPE: protein
16     OS-08-686-417-3

```

Query Match	16.0%;	Score 274.5;	DB 2;	Length 419;
Best Local Similarity	45.9%;	Pred. No. 2.4e-20;		
Matches 62;	Conservative 15;	Mismatches 23;	Indels 35;	Gaps 5;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:57:22 ; Search time 20.55 Seconds

(without alignments)
380.361 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGVQHRHASPEDIK.....EAERGVREHVDQRSIDRT 330

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 188963 seqs, 23686106 residues

Total number of hits satisfying chosen parameters: 188963

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A_Geneseq_36:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696	98.8	330	1 W94066	Human DnaJ-like pr
2	291	17.0	369	1 W98475	H. pylori GHPO 885
3	274.5	16.0	419	1 W00936	Maize DnaJ-related
4	274.5	16.0	419	1 W83397	Maize DnaJ clone p
5	272	15.9	223	1 W95710	Homo sapiens fetal
6	259.5	15.1	352	1 W22358	S. pneumoniae N-te
7	257	15.0	358	1 W94065	Human DnaJ-like pr
8	240	14.0	415	1 Y07061	Renal cancer assoc
9	232.5	13.5	438	1 W59132	Mus musculus Tubb 1
10	227.5	13.3	223	1 R90680	Mouse cysteine str
11	218.5	12.7	504	1 W36140	Bovine P58 protein
12	217	12.6	288	1 W98446	H. pylori GHPO 542
13	216	12.6	275	1 Y11104	H. pylori ORF 05ce
14	216	12.6	278	1 Y11105	H. pylori ORF hp7e
15	176	10.3	159	1 W44076	Human secreted pro
16	175	10.3	159	1 W27640	Secreted protein A
17	127.5	7.4	241	1 W87996	A human MCG18 prot
18	123.5	7.2	1073	1 P60569	Sequence of the Ri
19	122.5	7.1	265	1 W74905	Human secreted pro
20	114.5	6.7	621	1 Y07029	Breast cancer asso
21	104.5	6.1	320	1 W55828	Human heterogeneous
22	103.5	6.0	341	1 W26553	Human heterogeneous
23	103.5	6.0	353	1 W26552	Human heterogeneous
24	103.5	6.0	353	1 W54362	Heterogeneous nucl
25	103.5	6.0	353	1 W50921	Amino acid sequenc
26	103.5	6.0	433	1 W30256	Zuotin. Membranes
27	100.5	5.9	537	1 R73991	Bovine oviduct spe
28	100	5.8	668	1 W55483	H. pylori ORF 14ap
29	100	5.8	677	1 W55328	H. pylori ORF hp3e
30	98	5.7	707	1 R79912	Human nucleolin. A
31	98	5.7	707	1 W84052	Human V3 loop HIV
32	96.5	5.6	643	1 R60020	Fibronectin. New h
33	93.5	5.4	955	1 R42088	Human p50 or KBP1
34	93.5	5.4	955	1 R42234	Human p50 protein

35	93.5	5.4	955	1 R42235	Human p50 protein
36	93.5	5.4	955	1 R42236	Human p50 protein
37	93.5	5.4	955	1 R42237	Human p50 protein
38	93.5	5.4	955	1 R42238	Human p50 protein
39	93.5	5.4	955	1 R42239	Human p50 protein
40	93.5	5.4	955	1 R42240	Human p50 protein
41	93.5	5.4	955	1 R42241	Human p50 protein
42	93.5	5.4	955	1 R42242	Human p50 protein
43	93.5	5.4	955	1 R42243	Human p50 protein
44	93.5	5.4	955	1 R42244	Human p50 protein
45	93.5	5.4	955	1 R42245	Human p50 protein

ALIGNMENTS

RESULT	ID	Sequence	Score	Query Match	Length	DB ID	Description
1	W94066	W94066 standard; protein; 330 AA.	1696	98.8	330	1 W94066	Human DnaJ-like pr
AC	DT	09-APR-1999 (first entry)	291	17.0	369	1 W98475	H. pylori GHPO 885
DE	DE	Human DnaJ-like protein; HSPJ2.	274.5	16.0	419	1 W00936	Maize DnaJ-related
KW	KW	DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukemia; immune disorder; inflammation; tissue damage; diabetes; wound healing; chromosome mapping.	274.5	16.0	419	1 W83397	Maize DnaJ clone p
OS	OS	Homo sapiens.	272	15.9	223	1 W95710	Homo sapiens fetal
FT	FT	Key	259.5	15.1	352	1 W22358	S. pneumoniae N-te
FT	FT	Misc_difference 239	257	15.0	358	1 W94065	Human DnaJ-like pr
FT	FT	/label= unknown	240	14.0	415	1 Y07061	Renal cancer assoc
FT	FT	/note= "encoded by GST"	232.5	13.5	438	1 W59132	Mus musculus Tubb 1
FT	FT	/label= unknown	227.5	13.3	223	1 R90680	Mouse cysteine str
FT	FT	/note= "encoded by NCC"	218.5	12.7	504	1 W36140	Bovine P58 protein
FT	FT	/label= unknown	217	12.6	288	1 W98446	H. pylori GHPO 542
FT	FT	/note= "encoded by NNC"	216	12.6	275	1 Y11104	H. pylori ORF 05ce
FT	FT	/label= unknown	216	12.6	278	1 Y11105	H. pylori ORF hp7e
FT	FT	/note= "encoded by NCA"	176	10.3	159	1 W44076	Human secreted pro
FT	FT	/label= unknown	175	10.3	159	1 W27640	Secreted protein A
FT	FT	/note= "encoded by TTN"	127.5	7.4	241	1 W87996	A human MCG18 prot
FT	FT	/label= unknown	123.5	7.2	1073	1 P60569	Sequence of the Ri
FT	FT	/note= "encoded by NAA"	122.5	7.1	265	1 W74905	Human secreted pro
FT	FT	/label= unknown	114.5	6.7	621	1 Y07029	Breast cancer asso
FT	FT	/note= "encoded by NAA"	104.5	6.1	320	1 W55828	Human heterogeneous
FT	FT	/label= unknown	103.5	6.0	341	1 W26553	Human heterogeneous
FT	FT	/note= "encoded by NGC"	103.5	6.0	353	1 W26552	Human heterogeneous
FT	FT	/label= unknown	103.5	6.0	353	1 W54362	Heterogeneous nucl
FT	FT	/note= "encoded by NAA"	103.5	6.0	353	1 W50921	Amino acid sequenc
FT	FT	/label= unknown	103.5	6.0	433	1 W30256	Zuotin. Membranes
FT	FT	/note= "encoded by NGC"	100.5	5.9	537	1 R73991	Bovine oviduct spe
FT	FT	/label= unknown	100	5.8	668	1 W55483	H. pylori ORF 14ap
FT	FT	/note= "encoded by NAA"	100	5.8	677	1 W55328	H. pylori ORF hp3e
FT	FT	/label= unknown	98	5.7	707	1 R79912	Human nucleolin. A
FT	FT	/note= "encoded by NAA"	98	5.7	707	1 W84052	Human V3 loop HIV
FT	FT	/label= unknown	96.5	5.6	643	1 R60020	Fibronectin. New h
FT	FT	/note= "encoded by NAA"	93.5	5.4	955	1 R42088	Human p50 or KBP1
FT	FT	/label= unknown	93.5	5.4	955	1 R42234	Human p50 protein

CC production of recombinant proteins. Recombinant HSP71 and HSP72 are used
CC to raise Ab, therapeutically and to screen for specific binding agents.
CC Antagonists are used to treat or prevent a wide variety of solid cancers,
CC leukaemia and lymphoma; immune disorders (typical of many disclosed are
CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
CC and parasitic infections) and inflammation. Agonists may be used to treat
CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound
CC healing, and may be expressed from a vector containing the nucleic acids.
CC Fragments of the nucleic acids are used as primers and probes for
CC detecting and quantifying the HSP7-encoding nucleic acid in usual
CC hybridisation and/or amplification assays, therapeutically as antisense,
CC triplex-forming or ribozyme molecules, and for chromosome mapping.
SQ Sequence 330 AA;

Query Match 98.8%; Score 1696; DB 1; Length 330;
Best Local Similarity 100.0%; Pred. No. 1.2e-165;
Matches 330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDVYEVGVORHASPEDIKKAYKALKMHPDKNPENKEAEKFKQVAEAYVLSDAK 60
DB 1 MDVYEVGVORHASPEDIKKAYKALKMHPDKNPENKEAEKFKQVAEAYVLSDAK 60
QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFEGGADPFSFDFEDPFE 120
DB 61 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFEGGADPFSFDFEDPFE 120
QY 121 DFGNRRGPRGSRGSGSFSAFSGFSGFSSPDGTFTSGTSGHGLTFSSTSF 180
DB 121 DFGNRRGPRGSRGSGSFSAFSGFSGFSSPDGTFTSGTSGHGLTFSSTSF 180
QY 181 GSGMGNFKISTSTKWNKRKITTKRIVENGGERVEEDGOLKLTINGVADDLAXE 240
DB 181 GSGMGNFKISTSTKWNKRKITTKRIVENGGERVEEDGOLKLTINGVADDLAXE 240
QY 241 ERMRRGONVLPAPAGLRPPKPPRPSASLRHXPCHLSKEGEEDRPWAPXKXPLASXAG 300
DB 241 ERMRRGONVLPAPAGLRPPKPPRPSASLRHXPCHLSKEGEEDRPWAPXKXPLASXAG 300
QY 301 XHGXKRMXAERGVVEEVDORQSLDR 330
DB 301 XHGXKRMXAERGVVEEVDORQSLDR 330

RESULT 2
W98475
ID W98475 standard; Protein: 369 AA.

AC W98475; 31-MAR-1999 (first entry)
DT H. pylori GHPD 885 protein.
DE GHPD protein; Helicobacter infection; gastroduodenal disease; gastritis;
KW peptic ulcer disease.
OS Helicobacter pylori.
PN W09843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; US-902615.
PR 01-APR-1997; US-833457.
PR 24-JUN-1997; US-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PI (INNR) MERIEUX ORAVAX PASTEUR MERIEUX SERONS.
PI A1-garawi A, Kleuthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-542293/46.
DR N-PSDB: X14194.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 853-855; 2054pp; English.
CC This sequence represents a Helicobacter pylori GHPD protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with

CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 369 AA;

Query Match 17.0%; Score 291; DB 1; Length 369;
Best Local Similarity 32.0%; Pred. No. 1e-21;
Matches 93; Conservative 32; Mismatches 80; Indels 86; Gaps 12;

QY 2 VDYEVGVORHASPEDIKKAYKALKMHPDKNPENKEAEKFKQVAEAYVLSDAK 61
DB 3 IAYTELEFVKHSHNOETIKSYRKALKYIPDNAGDK-EAEKFKLINEAYVLSDEK 61
QY 62 RDYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDVREFEGGADPFSFDFEDP--- 119
DB 62 RALYDRYKKGKLNQAGASGDFS-----DEFDLGSF 93
QY 119 FEDFGNRRGPRGSRGSGT-----SFSFASGFP-----SFGSGFSSPD-TGF 161
DB 94 FEDAFG--FGAKSKKQKSSADYVLOTLELSEKAVFGCKTKIKVQYVSCSCGTGA 151
QY 162 -----TSGSLGSG-----GLTSFSTSGSGGNGFNSISTSTKWNKRKITTKRIY 209
DB 152 KDKALEFCQKQNGQGVFMNGFMSFAQTC--CAGCGKRIYVTPQACKGKTYILK--- 207
QY 210 ENGQERVEVEDGOLKLTINGVADDLA-----XEMRRGONVLPAP 253
DB 207 -----DEIDALIPGIDDONRMVLKNKGNEYKRGKGLYLEAQ 246

RESULT 3

W00936
ID W00936 standard; Protein: 419 AA.

AC W00936; 11-NOV-1997 (first entry)
DT Maize DnaJ-related protein.
DE DnaJ-related protein; chaperone; protein folding; DNA replication;
KW translation; peptide translocation; zmdJ1; promoter; maize;
KW transgenic plant; insecticide; antifungal; fungicide;
KW crop protection.
OS Zea mays L. cv. B73.
PN W09705260-A2.
PD 13-FEB-1997.
PF 12-JUL-1996; U11676.
PR 26-JUL-1995; US-001522.
PA (PION-) PIONEER HI-BRED INT INC.
PI Barbour E, Baszczynski C, Horowitz J, Rosichan JL;
DR WPI: 97-145697/13.
DR N-PSDB: T84330.
PT Tobacco DnaJ-related gene transcription/translation regulatory
PT sequence, zmdJ1 - is intermediate between constitutive and tissue
PT specific promoters, partic. for control of antifungal and
PT insecticide genes
PS Example 1; Fig 2A-B; 26pp; English.
CC This polypeptide sequence is encoded by the maize DnaJ-related
CC gene (see T84330). DnaJ-related proteins assist in chaperone-
CC mediated protein folding and provide cell viability at high
CC temperatures. They are also involved in DNA replication,
CC translation and peptide translocation across intracellular
CC membranes. Due to this wide range of functions, DnaJ has a wide
CC range of effectiveness and the gene's promoter sequence, zmdJ1
CC (see T84330), is effective in a wide range of tissues. The
CC claimed zmdJ1 promoter can be used in claimed methods for control
CC of antifungal or insecticidal genes in transgenic plants.
SQ Sequence 419 AA;

Query Match 16.0%; Score 274.5; DB 1; Length 419;
Best Local Similarity 45.9%; Pred. No. 6.1e-20;
Matches 62; Conservative 15; Mismatches 23; Indels 35; Gaps 5;

PF 17-MAY-1996; CA0322.
 PR 07-JUN-1995; US-472534.
 PR 04-AUG-1995; US-001805.
 PA (IAFB-) IAF BIOVAC INC.
 PI Brodeur B, Hamel J, Martin D, Rioux C;
 DR WPI: 97-052328/05.
 DR N-PSDB; T73388.
 PT Streptococcal heat shock proteins and corresponding DNA sequences -
 PT used in the production of a vaccine to treating and preventing
 PT strain-specific Streptococcal infection
 PS Example 3: Page 93-94; 156pp; English.
 CC This amino acid sequence corresponds to the N-terminal portion of the
 CC Streptococcus pneumoniae DnaJ protein. The protein shows a high degree
 CC of identity (72% and 51%) to the DnaJ proteins from *Lactococcus lactis*
 CC and *E. coli* respectively. The sequence encoding this truncated protein
 CC was isolated on the same nucleotide sequence that encodes the
 CC S. pneumoniae heat shock protein 72 (HSP72; W2357). The nucleotide
 CC fragment was isolated from a HindIII-partially digested genomic DNA
 CC library using a fragment of the chimeric gene (T73392) corresponding to
 CC the sequence encoding the C-terminal 169 amino acids of HSP72. The HSP72
 CC protein and its fragment, or antibodies specific to HSP72, are used in
 CC pharmaceutical compositions, pref. a vaccine, for treating or preventing
 CC infection by S. pneumoniae or related bacteria in humans, e.g.
 CC S. pyogenes or S. agalactiae.
 SQ Sequence 352 AA;

Query Match 15.1%; Score 259.5; DB 1; Length 352;
 Best Local Similarity 32.7%; Pred. No. 1.6e-18;
 Matches 81; Conservative 30; Mismatches 82; Indels 55; Gaps 10;

QY 3 DYEVGVQVHNASPEDIKKAYRKALKMHPDKPENKEAEKFKOVAEAYEVLSDAKKR 62
 DB 5 EYDRLGVSNANADEIKKAYRKALKMHPDKPENKEAEKFKOVAEAYEVLSDAKKR 62
 QY 63 DIYDKYKGLNGGGGSHFDSPEFGFTFRNPDVFEFFEGG---RDP----- 110
 DB 63 AADYQAGANGANGFGAGGFGG-FNGAGGFGFEDITFSFGGSSSRPNAPRGGDDI 121
 QY 110 -----FSPDFEEDPFDFEGGNR-----GPRGSRNCTGSGFSAFSGFPFGSGFSS 156
 DB 122 QYRVNLTFF-----EAFETKEVKYHVRAGCRTGNGSAGKPGTSPYVCGRGHAGVYN 175
 QY 157 FDTGFTSFGSL-----GHGILTSFSTSGSGMGNGEKSTSTKMANGRKITTAK 206
 DB 176 VDTQ-TPLGMRQVTCVCHGRGKEIKYCTTCHGTGHEK-QAHSVHVKIPAG----- 228
 QY 207 RIVENGQOE 214
 DB 228 --VETGQQ 233

RESULT 7
 ID W94065 standard; Protein: 358 AA.
 AC W94065;
 DT 09-APR-1999 (first entry)
 DE Human DnaJ-like protein, HSPJ1.
 KW DnaJ-like protein; heat shock protein; HSPJ1; HSPJ2; cancer; leukaemia;
 KW immune disorder; inflammation; tissue damage; diabetes; wound healing;
 KW chromosome mapping.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Misc_difference 298
 FT /Label= unknown
 FT /note= "encoded by AAN"
 FT Misc_difference 299
 FT /Label= unknown
 FT /note= "encoded by TAN"
 PW W09855509-A2.
 PD 10-DEC-1998.
 PF 02-JUN-1998; U11182.
 PR 03-JUN-1997; US-868288.

PA (INCY-) INCYTE PHARM INC.
 PI Al-Young J, Bandman O, Lai P;
 DR WPI: 99-070259/06.
 DR N-PSDB; X06099.
 PT New nucleic acid encoding human DnaJ-like proteins - for diagnosis,
 PT treatment and prevention of cancer, immune disorders and
 PT inflammation
 PS Claim 1: Fig 1A-D; 73pp; English.
 CC This represents a human DnaJ-like protein, HSPJ1. The invention provides
 CC two human DnaJ-like proteins which are heat shock proteins J1 and J2
 CC (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host
 CC cells containing a vector comprising the nucleic acids are used for the
 CC production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used
 CC to raise Ab, therapeutically and to screen for specific binding agents.
 CC Antagonists are used to treat or prevent a wide variety of solid cancers,
 CC leukemia and lymphoma; immune disorders (typical of many disclosed are
 CC acquired immune deficiency syndrome, allergy, asthma, Crohn's disease,
 CC diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial
 CC and parasitic infections) and inflammation. Agonists may be used to treat
 CC or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart
 CC attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound
 CC healing, and may be expressed from a vector containing the nucleic acids.
 CC Fragments of the nucleic acids are used as primers and probes for
 CC detecting and quantifying the HSPJ-encoding nucleic acid in usual
 CC hybridisation and/or amplification assays, therapeutically as antisense,
 CC triplex-forming or ribozyme molecules, and for chromosome mapping.
 SQ Sequence 358 AA;

Query Match 15.0%; Score 257; DB 1; Length 358;
 Best Local Similarity 46.2%; Pred. No. 3e-18;
 Matches 60; Conservative 18; Mismatches 30; Indels 22; Gaps 5;

QY 3 DYEVGVQVHNASPEDIKKAYRKALKMHPDKPENKEAEKFKOVAEAYEVLSDAKKR 62
 DB 25 DYFKILGVRSASIKDIKAYRKALKMHPDKPENKEAEKFKOVAEAYEVLSDAKKR 83
 QY 63 DIYDKYKGLNGGGGSHFDSPEFGFTFRNPDVFEFFEGG---RDP----- 122
 DB 84 KQYDYTGEGGLK-DGHQSH-----GDTSHFFG---DFGFMGCTPRQO- 125
 QY 123 FGNRRGPRGS 132
 DB 125 --DRNIPRGS 132

RESULT 8
 ID Y07061 standard; Protein: 415 AA.
 AC Y07061;
 DT 02-JUL-1999 (first entry)
 DE Renal cancer associated antigen precursor sequence.
 KW Cancer associated antigen; diagnosis; research; treatment; human;
 KW breast cancer; colon cancer; gastric cancer; renal cancer; lung cancer;
 KW prostate cancer.
 OS Homo sapiens.
 FH Homo sapiens.
 PN W09904265-A2.
 PD 28-JAN-1999.
 PF 15-JUL-1998; U14679.
 PR 22-JUN-1998; U14679.
 PR 17-JUL-1997; US-896164.
 PR 10-OCT-1997; US-061599.
 PR 10-OCT-1997; US-061765.
 PR 10-OCT-1997; US-948705.
 PR 11-OCT-1997; GB-021697.
 PA (LUDW-) LUDWIG INST CANCER RES.
 PI Chen Y, Cout I, Gure A, Ohare M, Odata Y, Old LJ,
 PI Pfreundschuh M, Sahin U, Scanlan MJ, Stockert E,
 PI Tureci O;
 DR WPI: 99-132448/11.
 PT New isolated cancer associated nucleic acids and polypeptides -
 PT isolated using sera from cancer patients, used to develop products
 PT for the diagnosis, monitoring or treatment of cancers

PS Disclosure: Page 456-457; 787P; English.
 CC The invention relates to a method for diagnosing a disorder characterised
 CC by expression of a human cancer associated antigen precursor coded for by
 CC a nucleic acid molecule (NAM). The method comprises: (a) contacting a
 CC biological sample isolated from a subject with an agent that specifically
 CC binds to the NAM, an expression product or a fragment of an expression
 CC product complexed with an HLA molecule; and (b) determining the
 CC interaction between the agent and the NAM or the expression product as a
 CC determination of the disorder. The products and methods can be used in
 CC the diagnosis, monitoring, research, or treatment of conditions
 CC characterised by the expression of various cancer associated antigens.
 CC The invention provides nucleic acid sequences and encoded polypeptides
 CC which are cancer associated antigen precursors expressed in human breast
 CC cancer, renal cancer, colon cancer, gastric cancer, prostate cancer and
 CC lung cancer.
 SQ Sequence 415 AA;

Query Match 14.0%; Score 240; DB 1; Length 415;
 Best Local Similarity 41.4%; Pred. No. 2.1e-16;
 Matches 55; Conservative 19; Mismatches 31; Indels 28; Gaps 5;

OY 5 YEVLGVORHASPEDIKAYRKALKMHPDKNPKNEKAEKFKQVAEAYEVLSDAKKRD 64
 ID W59132 standard; Protein; 438 AA.
 AC W59132;
 DT 11-SEP-1998 (first entry)
 DE Mus musculus Tubb Interactor (MTI-3) protein.
 KM serine protease; tub interactor; treatment; obesity; cachexia;
 KM anorexia nervosa; diabetes; cell cycle progression; apoptosis;
 KM neurodegenerative disease; Alzheimer's disease; drug screening;
 KM Parkinson's disease; Huntington's chorea; detection; diagnosis;
 KM amyotrophic lateral sclerosis; spinocerebellar degeneration.
 OS Mus musculus.
 PN W09812302-A1.
 PD 26-MAR-1998.
 PF 05-SEP-1997; U15627.
 PR 21-JUL-1997; US-897340.
 PR 17-SEP-1996; US-715032.
 PA (MILL-) MILLENNIUM PHARM INC.
 PI Errada PR, Gimeno CJ;
 DR WPI: 98-217246/19.
 DR N-PSDB; V11858.
 PT Tubb Interactor genes - used to develop products for the treatment
 PT of obesity, cachexia, anorexia nervosa or related disorders e.g.
 PT diabetes
 PT Claim 28; Fig 4; 120P; English.
 CC The sequence is that encoding the Tubb Interactor protein (MTI-3)
 CC which is a putative serine protease. TI genes function
 CC in biochemical pathways involved in weight control and
 CC related disorders. The products can be used for treating
 CC weight disorders, e.g. obesity, cachexia or anorexia nervosa,
 CC or a related disorder such as diabetes. The products can
 CC also be used to modulate cell cycle progression and apoptosis.
 CC They can be used for treating neurodegenerative diseases
 CC which are characterised by apoptosis, including Alzheimer's
 CC disease, Parkinson's disease, Huntington's chorea, amyotrophic
 CC lateral sclerosis or spinocerebellar degenerations. The
 CC products can also be used for detection, diagnosis and
 CC drug screening.

SQ Sequence 438 AA;

Query Match 13.5%; Score 232.5; DB 1; Length 438;
 Best Local Similarity 36.2%; Pred. No. 1.3e-15;
 Matches 54; Conservative 26; Mismatches 26; Indels 43; Gaps 5;

OY 3 DYEVLTGQRHASPEDIKAYRKALKMHPDKNPKNEKAEKFKQVAEAYEVLSD 58
 ID W59132 standard; Protein; 438 AA.
 AC W59132;
 DT 07-APR-1996 (first entry)
 DE Mouse cysteine string protein.
 KM Mouse; cysteine string protein; anergy; T-lymphocyte;
 KM differential display; cDNA; reverse transcription; probe;
 KM polymerase chain reaction; cloning; gel electrophoresis;
 KM antibody; diagnostic; immunoassay; autoimmune disease;
 KM organ transplantation; vaccine; drug screening.
 OS Mus musculus.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LEHMAN STANFORD JUNIOR.
 PI Bloom D, Fachman G;
 DR WPI: 96-068884/07.
 DR N-PSDB; T11786.
 PT Identifying genes associated with T-cell anergy, and the proteins
 PT encoded by them - useful to elucidate the anergic state, and
 PT identify agents associated with induction of anergy
 PS Disclosure; Fig 2; 31P; English.
 CC The sequence represents a mouse cysteine string protein encoded by
 CC a gene isolated from mouse T-cell clones 11.3.7 and 12.2 by a new
 CC procedure for identification of genes associated with anergy. The
 CC protein is expressed in resting T-cells but not in anergic T-cells.
 CC To isolate proteins of this type, differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.

OY 116 EDPEDFFGNRRGPRGSRNRTGSGFFSAF 144
 ID W59132 standard; Protein; 438 AA.
 AC W59132;
 DT 07-APR-1996 (first entry)
 DE Mouse cysteine string protein.
 KM Mouse; cysteine string protein; anergy; T-lymphocyte;
 KM differential display; cDNA; reverse transcription; probe;
 KM polymerase chain reaction; cloning; gel electrophoresis;
 KM antibody; diagnostic; immunoassay; autoimmune disease;
 KM organ transplantation; vaccine; drug screening.
 OS Mus musculus.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LEHMAN STANFORD JUNIOR.
 PI Bloom D, Fachman G;
 DR WPI: 96-068884/07.
 DR N-PSDB; T11786.
 PT Identifying genes associated with T-cell anergy, and the proteins
 PT encoded by them - useful to elucidate the anergic state, and
 PT identify agents associated with induction of anergy
 PS Disclosure; Fig 2; 31P; English.
 CC The sequence represents a mouse cysteine string protein encoded by
 CC a gene isolated from mouse T-cell clones 11.3.7 and 12.2 by a new
 CC procedure for identification of genes associated with anergy. The
 CC protein is expressed in resting T-cells but not in anergic T-cells.
 CC To isolate proteins of this type, differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.

Query Match 13.3%; Score 227.5; DB 1; Length 223;
 Best Local Similarity 60.9%; Pred. No. 1.6e-15;
 Matches 54; Conservative 26; Mismatches 26; Indels 43; Gaps 5;

OS Helicobacter pylori.
PN MO9843478-A1.
PD 08-OCT-1998.
PF 01-APR-1998; U06371.
PR 29-JUL-1997; U5-902615.
PR 01-APR-1997; U5-833457.
PR 24-JUN-1997; U5-881227.
PA (HUMA-) HUMAN GENOME SCI INC.
PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
PI Al-garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
DR WPI: 98-342293/46.
DR N-PSDB; X14165.
PT New isolated Helicobacter polynucleotides - used to develop products
PT for the diagnosis, prevention and treatment of Helicobacter
PT infections and gastrointestinal diseases
PS Claim 8; Page 773-774; 2054pp; Eng11sh.
CC This sequence represents a Helicobacter pylori GHPD protein of the
CC invention. The polypeptides can be used for preventing or treating
CC Helicobacter infections, and gastroduodenal diseases associated with
CC these infections, including acute, chronic, and atrophic gastritis, and
CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
CC used for the production of antibodies. The products can also be used for
CC detection and diagnosis.
SQ Sequence 288 AA;

Query Match 12.6%; Score 217; DB 1; Length 288;
Best Local Similarity 27.9%; Pred No. 2.7e-14;
Matches 69; Conservative 34; Mismatches 84; Indels 60; Gaps 9

QY 5 YEVLGVORHASPEDIKKAYRLALKWHDPKNPENKEAEKRFKQVAEAYEVLDAKKRDI 64
Db 6 YÖTLUNSENA5ODEIKKSYRLARQYHPDLN--KTKAEKEKFEKIWAAYEILSDDEKKRQ 63
QY 65 YDKYKEGELNGGGGGSHD5PFEFGFTFRNDDVYREFFEGGRDPRF5PFEDPFEDPF 124
Db 64 YDÖFED-----NMFGQN-----FSD-PA 81
QY 125 NNRGPGSRKSGTGF5FFSAFSGF5SGF5SFDIGFT5F5GLGHG---GLTSF5ST5F 180
Db 82 RSRGP-----SELDLIL5IFGKG5F5Q5R5F5G5F5G5F5NS5N5PA5EM5DIT5AL5NV5 137
QY 181 GSGGNGKNSKIST5TK5M5N5G5K5IT5TK5RI--VEN5G5R5VE5ED5G5L5IT5NG5V5DD5AL 238
Db 138 LDTLTLGNKRKQV5-----INN5TF5SL5K5IP5IG5E5G--EK5IR5V5RN5K5G5TK5RT5RD5ILL5E5IH 191
QY 239 XEERMR 245
Db 192 EDEMYR 198

RESULT 13
Y11104
AC Y11104 standard; Protein; 275 AA.
DF 08-JUN-1999 (first entry)
H. pylori ORF 05ce10910_23712780.c1.4 cytoplasmic protein.
KW vaccine; probe; diagnostic; ORF; cell envelope protein.
NS secreted protein; cytoplasmic protein; cellular protein.
OS Helicobacter pylori.
PN MO9824475-A1.
PD 11-JUN-1998.
PF 05-DEC-1997; U22104.
PR 14-JUL-1997; U5-891928.
PR 05-DEC-1996; U5-759625.
PR 25-MAR-1997; U5-823745.
PA (ASTR) ASTRA AB.
PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
DR WPI: 98-33051/29.
DR N-PSDB; X30633.
PT New isolated Helicobacter pylori nucleic acids - used to develop
PT products for the diagnosis, prevention and treatment of infection by
PT H. pylori and other Helicobacter species

PS Claims 37, 41; Page 270-271; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 275 AA;

Query Match 12.6%; Score 216; DB 1; Length 275;
 Best Local Similarity 29.2%; Pred. No. 3.2e-14;
 Matches 66; Conservative 30; Mismatches 70; Indels 60; Gaps 9;

QY 5 YEVLGVRHASPEDIKKAYRKALKMHPDKNPENKEAEKRFQVAAVEYLSDAKRD 64
 DB 6 YQTLNVEENASQDEIKSYRLAQQYHPDLN--KTKAEAEKFEINAAVEYLSDEKRRQ 63
 QY 65 YDKYKREGLNGGGGGSHFDSPEFGFTFRNPDDVREFRFGGRDPFSFDEPDEDFPG 124
 DB 64 YDQFGD-----NMFGGN-----FSD-FA 81
 QY 125 NRRGPRGSRGRTGTSFSAFGSPFGSGFSDPTGFTSGSLGHC---GLTFSSTSF 180
 DB 82 RSRGP-----SEDDLDLISLTFKGGFSQRFSGFSGFNFSNAPENDVTAALNVSV 137
 QY 181 GSGMGNFKSIISTKMWNGRKITTKRI--VENGOEVEVEEDGOL 224
 DB 138 LDTLLGNKKQVS-----VNNETFSLKIPIGVEEG-EKIRVRNKGKM 177

RESULT 14

ID Y11105 standard; Protein; 278 AA.
 AC Y11105;
 DT 08-JUN-1999 (first entry)
 DE H. pylori ORF hp7e10192_23712780.f2.5 cytoplasmic protein.
 KW Vaccine; probe; diagnostic; ORF; cell envelope protein;
 KM secreted protein; cytoplasmic protein; cellular protein.
 OS Helicobacter pylori.
 PN WO9824475-A1.
 PD 11-JUN-1998.
 PF 05-DEC-1997; U22104.
 PR 14-JUL-1997; US-891928.
 PR 05-DEC-1996; US-759625.
 PR 25-MAR-1997; US-823745.
 PA (ASTR) ASTRA AB.
 PI Alm RA, Castriotta LM, Doig PC, Kabok Z, Smith D;
 DR WPI: 98-33051/29.
 DR N-PSDB: X30634.
 PT New isolated Helicobacter pylori nucleic acids - used to develop
 PT products for the diagnosis, prevention and treatment of infection by
 PT H. pylori and other Helicobacter species
 PS Claims 37, 41; Page 271-272; 339pp; English.
 CC Recombinant or substantially pure preparations of H. pylori polypeptides
 CC are disclosed, together with the nucleic acids encoding them. In all,
 CC 97 ORFs are shown. The proteins are variously cell envelope proteins,
 CC cytoplasmic proteins, secreted proteins or other cellular proteins.
 CC Vaccines containing the nucleic acids or proteins are claimed, as are
 CC probes containing at least 8 nucleotides from the nucleic acid
 CC sequences. The vaccines are useful for treating or reducing the risk of
 CC H. pylori infections, and the probes can be used diagnostically for
 CC detecting the presence of Helicobacter in a sample. The products are
 CC also of use in screening for compounds having the ability to interfere
 CC with the H. pylori life cycle or to inhibit H. pylori infection.
 SO Sequence 278 AA;

Query Match 12.6%; Score 216; DB 1; Length 278;
 Best Local Similarity 29.2%; Pred. No. 3.3e-14;
 Matches 66; Conservative 30; Mismatches 70; Indels 60; Gaps 9;

QY 5 YEVLGVRHASPEDIKKAYRKALKMHPDKNPENKEAEKRFQVAAVEYLSDAKRD 64
 DB 6 YQTLNVEENASQDEIKSYRLAQQYHPDLN--KTKAEAEKFEINAAVEYLSDEKRRQ 63
 QY 65 YDKYKREGLNGGGGGSHFDSPEFGFTFRNPDDVREFRFGGRDPFSFDEPDEDFPG 124
 DB 64 YDQFGD-----NMFGGN-----FSD-FA 81
 QY 125 NRRGPRGSRGRTGTSFSAFGSPFGSGFSDPTGFTSGSLGHC---GLTFSSTSF 180
 DB 82 RSRGP-----SEDDLDLISLTFKGGFSQRFSGFSGFNFSNAPENDVTAALNVSV 137
 QY 181 GSGMGNFKSIISTKMWNGRKITTKRI--VENGOEVEVEEDGOL 224
 DB 138 LDTLLGNKKQVS-----VNNETFSLKIPIGVEEG-EKIRVRNKGKM 177

RESULT 15

ID W44076 standard; Protein; 159 AA.
 AC W44076;
 DT 12-MAY-1998 (first entry)
 DE Human secreted protein Am610.
 KW Human; secreted protein; ATCC 98026; cytokine; immunomodulation;
 OS cell proliferation; differentiation; regulation.
 PH Homo sapiens.
 FT location/Qualifiers
 FT Misc.-difference 82
 FT /label= Unspecified
 FT /note= "encoded by AGN"
 FT Misc.-difference 87
 FT /label= Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc.-difference 96
 FT /label= Unspecified
 FT /note= "encoded by TAG a stop codon"
 FT Misc.-difference 102
 FT /label= Unspecified
 FT /note= "encoded by RGR"
 FT Misc.-difference 103
 FT /label= Unspecified
 FT /note= "encoded by RGT"
 FT Misc.-difference 106
 FT /label= Unspecified
 FT /note= "encoded by RAG"
 FT Misc.-difference 109
 FT /label= Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc.-difference 110
 FT /label= Unspecified
 FT /note= "encoded by GYR"
 FT Misc.-difference 111
 FT /label= Unspecified
 FT /note= "encoded by NKC"
 FT Misc.-difference 112
 FT /label= Unspecified
 FT /note= "encoded by NKT"
 FT Misc.-difference 113
 FT /label= Unspecified
 FT /note= "encoded by TMN"
 FT Misc.-difference 114
 FT /label= Unspecified
 FT /note= "encoded by YTT"
 FT Misc.-difference 115
 FT /label= Unspecified
 FT /note= "encoded by YAA"
 FT Misc.-difference 116
 FT /label= Unspecified
 FT /note= "encoded by YTT"

FT Misc_difference 117 /label= Unspecified
 FT /note= "encoded by TSA"
 FT Misc_difference 118 /label= Unspecified
 FT /note= "encoded by TGA a stop codon"
 FT Misc_difference 121 /label= Unspecified
 FT /note= "encoded by TAA a stop codon"
 FT Misc_difference 128 /label= Unspecified
 FT /note= "encoded by TYN"
 FT Misc_difference 129 /label= Unspecified
 FT /note= "encoded by ARR"
 FT Misc_difference 130 /label= Unspecified
 FT /note= "encoded by CVA"
 FT Misc_difference 134 /label= Unspecified
 FT /note= "encoded by AKC"
 FT Misc_difference 135 /label= Unspecified
 FT /note= "encoded by YAA"
 FT Misc_difference 136 /label= Unspecified
 FT /note= "encoded by NAA"
 FT Misc_difference 137 /label= Unspecified
 FT /note= "encoded by YKR"
 FT Misc_difference 139 /label= Unspecified
 FT /note= "encoded by GRP"
 FT Misc_difference 140 /label= Unspecified
 FT /note= "encoded by RWY"
 FT Misc_difference 141 /label= Unspecified
 FT /note= "encoded by CAW"
 FT Misc_difference 142 /label= Unspecified
 FT /note= "encoded by WYC"
 FT Misc_difference 143 /label= Unspecified
 FT /note= "encoded by CNN"
 FT Misc_difference 146 /label= Unspecified
 FT /note= "encoded by NNN"
 FT Misc_difference 147 /label= Unspecified
 FT /note= "encoded by WKG"
 FT Misc_difference 149 /label= Unspecified
 FT /note= "encoded by SYC"
 FT Misc_difference 151 /label= Unspecified
 FT /note= "encoded by GNG"
 FT Misc_difference 155 /label= Unspecified
 FT /note= "encoded by TGN"
 PN W09739123-A2.
 PD 23-OCT-1997.
 PF 14-APR-1997; U06139.
 PR 18-APR-1996; US-634325.
 RA (GENY) GENETICS INST INC.
 PI Jacobs K, Lavallie ER, McCoy JM, Merberg D, Racle LA,
 PI Spaulding V;
 DR WPI: 97-526460/48.
 DR N-PSDB; V02136.
 DR New secreted proteins encoded clones present in ATCC 98026 -
 PT possibly having cytokine, cell proliferation/differentiation
 PT regulating, immunomodulating and many other activities
 PS Clatm 12; Page 78-79; 139pp. English.

CC The present sequence represents a novel human secreted protein deposited
 CC under accession number ATCC 98026. The secreted protein can be used to
 CC determine biological activity, to raise antibodies, as tissue markers,
 CC to isolate cognate ligands or receptors, to identify agents that
 CC modulate their interactions and as nutritional supplements. It may also
 CC have a very wide range of biological activities although no evidence
 CC for any is provided in the specification. Typical of these are cytokine,
 CC cell proliferation/differentiation modulating activity or induction of
 CC other cytokines; immunostimulating/immunosuppressant activities (e.g.
 CC for treating human immunodeficiency virus infection, cancer, autoimmune
 CC diseases and allergy); regulation of haematopoiesis (e.g. for treating
 CC anaemia or as adjunct to chemotherapy); stimulation of growth of bone,
 CC cartilage, tendons, ligaments and/or nerves (e.g. for treating wounds,
 CC periodontal disease, neurological diseases stroke, fibrosis); inhibition
 CC or stimulation of follicle stimulating hormone (for control of
 CC fertility); chemotactic and chemokinetic activities (e.g. for treating
 CC infections, tumours); haemostatic or thrombolytic activity (e.g. for
 CC treating haemophilia, cardiac infarction etc.); anti-inflammatory
 CC activity (e.g. for treating septic shock, Crohn's disease); as
 CC antimicrobials; for treating psoriasis or other hyperproliferative
 CC disease; for regulation of metabolism, behaviour, and many others. Also
 CC contemplated is the use of the corresponding nucleic acid in gene
 CC therapy procedures.
 CC Sequence 159 AA:

Query Match 10.3%; Score 176; DB 1; Length 159;
 Best Local Similarity 56.2%; Pred. No. 1.8e-10;
 Matches 36; Conservative 10; Mismatches 14; Indels 4; Gaps 2;
 QY 4 YVEVLGVORHASPEDIKKAYRKLALMKWHPDKNPENKEAEERFKQYAEYVLSDAK--K 61
 Db 27 YVDILGVPRKSASBRQIKKAFHKLAMKYHPDKN--KSPDAENAFRELAERYETLSDAKXK 84
 QY 62 RDLY 65
 Db 85 RYXY 88

Search completed: June 30, 2000, 09:57:24
 Job time: 2725 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 13:32:40 ; Search time 49.41 Seconds
(without alignments)
3498.889 Million cell updates/sec

Title: US-09-501-714-4

Perfect score: 1330

Sequence: 1 CGNAGAGAGGNAAGGAAG.....TGACGGCAGCGGTGGCGGG 1330

Scoring table:

IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 6492525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: /cgn2_patents.NA:*
2: /cgn2_6/ptodata/2/1na/5A.COMB.seq:*
3: /cgn2_6/ptodata/2/1na/5B.COMB.seq:*
4: /cgn2_6/ptodata/2/1na/5D.COMB.seq:*
5: /cgn2_6/ptodata/2/1na/6.COMB.seq:*
6: /cgn2_6/ptodata/2/1na/PCITUS.COMB.seq:*
7: /cgn2_6/ptodata/2/1na/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1307.2	98.3	1330	4	Sequence 4, Appl1
2	1307.2	98.3	1330	5	Sequence 4, Appl1
3	1307.2	98.3	1330	5	Sequence 4, Appl1
4	91.4	6.9	1376	4	Sequence 2, Appl1
5	91.4	6.9	1376	5	Sequence 2, Appl1
6	91.4	6.9	1376	5	Sequence 2, Appl1
7	75.6	5.7	672	2	Sequence 6, Appl1
8	73.4	5.5	2349	4	Sequence 2, Appl1
9	68.2	5.1	4320	4	Sequence 4, Appl1
10	66.4	5.0	1756	4	Sequence 3, Appl1
11	60	4.5	1515	2	Sequence 1, Appl1
12	60	4.5	1687	1	Sequence 14, Appl1
13	55.8	4.2	1700	1	Sequence 14, Appl1
14	55.2	4.2	1700	1	Sequence 14, Appl1
15	54	4.1	3748	3	Sequence 2, Appl1
16	34.4	2.6	13987	3	Sequence 13, Appl1
17	34.4	2.6	44377	3	Sequence 7, Appl1
18	34.4	2.6	44377	3	Sequence 7, Appl1
19	34.2	2.6	1704	2	Sequence 5, Appl1
20	34.2	2.6	1704	2	Sequence 5, Appl1
21	34.2	2.6	1908	2	Sequence 8, Appl1
22	34.2	2.6	1908	2	Sequence 8, Appl1
23	34	2.6	1787	3	Sequence 8, Appl1
24	34	2.6	3211	4	Sequence 8, Appl1
25	34	2.6	3901	4	Sequence 8, Appl1
26	33.8	2.5	1017	6	Sequence 21, Appl1
27	33.4	2.5	1785	5	Sequence 16, Appl1

28	33	2.5	1662	2	US-08-671-947-1	Sequence 1, Appl1
29	32.6	2.5	1644	1	US-07-903-047-7	Sequence 7, Appl1
30	32.6	2.5	3901	1	US-08-188-582-31	Sequence 31, Appl1
31	32.6	2.5	3901	1	US-08-646-715-31	Sequence 7, Appl1
32	32.6	2.5	44377	3	US-08-804-227C-7	Sequence 1, Appl1
33	32.6	2.5	44377	4	US-08-804-198-1	Sequence 1, Appl1
34	32.4	2.4	35081	3	US-08-752-760A-1	Sequence 1, Appl1
35	31.8	2.4	1055	2	US-08-618-464-1	Sequence 1, Appl1
36	31.8	2.4	1055	6	PCT-US95-04896-1	Sequence 1, Appl1
37	31.8	2.4	1820	1	US-08-173-508-7	Sequence 7, Appl1
38	31.8	2.4	1821	3	US-08-265-310-7	Sequence 7, Appl1
39	31.8	2.4	2312	1	US-07-736-178C-1	Sequence 1, Appl1
40	31.8	2.4	30001	1	US-08-125-468-1	Sequence 1, Appl1
41	31.8	2.4	30001	3	US-08-474-933-1	Sequence 1, Appl1
42	31.4	2.4	4852	1	US-07-853-913-3	Sequence 1, Appl1
43	31.2	2.3	370	3	US-08-332-766A-8	Sequence 8, Appl1
44	31.2	2.3	879	1	US-08-243-545-1	Sequence 1, Appl1
45	31.2	2.3	879	3	US-08-993-962-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-868-288A-4
Sequence 4, Application US/08868288A
Patent No. 5923567
GENERAL INFORMATION:
APPLICANT: Au-Young, Janice
APPLICANT: Lal, Preeti
APPLICANT: Bandman, Olga
TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/868,288A
FILING DATE: June 3, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0309 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1330 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: HMT2RAT01
CLONE: 260873
US-08-868-288A-4
Query Match 98.3%; Score 1307.2; DB 4; Length 1330;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1330; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	61	GGGCGNTGAGGAGATGTGGGGCCGTAACCGGCTCCCGCTCCCGGACCGGGCGCT	120
Db	61	GGNGCTNTAGGAGATGTGGGGCGGACCGCTCCCTCCCTCGCTCCCGCCACGGGCGCT	120
Qy	121	TCCTTCCTCGGAGCCATTCCAAACATCTCGTAACACATGTGTGATTAATCTGAAGTTCTA	180
Db	121	TCCTTCCTCGGAGCCATTCCAAACATCTCGTAACACATGTGTGATTAATCTGAAGTTCTA	180
Qy	181	GGGCTGCAGACATGTCCCTCACCCGAGATATTAAAAAGCATATCGGAACCTGGCACTG	240
Db	181	GGGCTGCAGACATGTCCCTCACCCGAGATATTAAAAAGCATATCGGAACCTGGCACTG	240
Qy	241	AAGTGCATCCAGATAAAATCTGAGAAATTAAGAACACAGAGAGAAAATTCAAGCAA	300
Db	241	AAGTGCATCCAGATAAAATCTGAGAAATTAAGAACACAGAGAGAAAATTCAAGCAA	300
Qy	301	GTAGCGGAGGATATGAAAGTCTGTGGATCTAAGAAACGGACATCTATGACAAATAT	360
Db	301	GTAGCGGAGGATATGAAAGTCTGTGGATCTAAGAAACGGACATCTATGACAAATAT	360
Qy	361	GGCAAAAGAGATTAATATGTGTGNGNGNGNGTGGAGTCAATTTAGACGTCACTTGA	420
Db	361	GGCAAAAGAGATTAATATGTGTGNGNGNGTGGAGTCAATTTAGACGTCACTTGA	420
Qy	421	TTTGGCTTCACATTCGGTAAACCAGATATGTCTTCAGGAAATTTTTGGTGAAGGAC	480
Db	421	TTTGGCTTCACATTCGGTAAACCAGATATGTCTTCAGGAAATTTTTGGTGAAGGAC	480
Qy	481	CCATTTTCATTTGACTCTTTGAGAACCCCTTTTGAGACTCTTTGGGAAATCGAAGGGT	540
Db	481	CCATTTTCATTTGACTCTTTGAGAACCCCTTTTGAGACTCTTTGGGAAATCGAAGGGT	540
Qy	541	CCCCGAGAGACAGAACGCCAGGAGCGGGGCGTTTTCTGCTGTAAGTGGATTTCCG	600
Db	541	CCCCGAGAGACAGAACGCCAGGAGCGGGGCGTTTTCTGCTGTAAGTGGATTTCCG	600
Qy	601	TCCTTTGGAGTGGATTTTCTCTTTTGATACAGATTTACTTCATTTGGGTACTAGGT	660
Db	601	TCCTTTGGAGTGGATTTTCTCTTTTGATACAGATTTACTTCATTTGGGTACTAGGT	660
Qy	661	CACGGGGGCTCACATTCATTCGTCACACGCAATTTGGTGTGGCATGGGCAAACTTC	720
Db	661	CACGGGGGCTCACATTCATTCGTCACACGCAATTTGGTGTGGCATGGGCAAACTTC	720
Qy	721	AAATGATATCACTTCACTAAATGTTAATGGCAGAAAATCACTCAAAAGAAAT	780
Db	721	AAATGATATCACTTCACTAAATGTTAATGGCAGAAAATCACTCAAAAGAAAT	780
Qy	781	GTGCGAAGCGGTCAAGAAAGTGAAGTGAAGAAAGATGGCAGTTAAAGTCTTAAACA	840
Db	781	GTGCGAAGCGGTCAAGAAAGTGAAGTGAAGAAAGATGGCAGTTAAAGTCTTAAACA	840
Qy	841	ATTAATGTGTGTGCGCAGAGAGATGCCCTGTGTAGAGAGCGCATGGCAGAGCCAGAAC	900
Db	841	ATTAATGTGTGTGCGCAGAGAGATGCCCTGTGTAGAGAGCGCATGGCAGAGCCAGAAC	900
Qy	901	GTCCGTGCCAGCCAGCCTGCCGGGCTCTCCAGCCGCCAGACCCCGCGGCTCTCGTTG	960
Db	901	GTCCGTGCCAGCCAGCCTGCCGGGCTCTCCAGCCGCCAGACCCCGCGGCTCTCGTTG	960
Qy	961	CTGAGACACAGNGCCTCATTTGTCTCTCTAAGAGAGAGCGGACAGACGACCTTGGGCA	1020
Db	961	CTGAGACACAGNGCCTCATTTGTCTCTCTAAGAGAGAGCGGACAGACGACCTTGGGCA	1020
Qy	1021	CCCGNGNCTGTGNNCCCCCTCGCTTCNCAGCAGGNTTNNAGAAGTNGCAAGAGATG	1080
Db	1021	CCCGNGNCTGTGNNCCCCCTCGCTTCNCAGCAGGNTTNNAGAAGTNGCAAGAGATG	1080
Qy	1081	NAACGAGAGAGAGAGAGAGTGCAGAAAGAAAGATGACCAAAAGCAATCACTATAC	1140
Db	1081	NAACGAGAGAGAGAGAGAGTGCAGAAAGAAAGATGACCAAAAGCAATCACTATAC	1140

QY	1141	CGAGATTAGGCACGCGGTGCACCCCCAGACGGTGGCGCTCCACCGTGTGGCATGGGG	1200
Db	1141	CGGACTTGAAGCAGCGCGTGCACCCCCAGACGGTGGCGCTCCACCGTGTGGCATGGGG	1200
QY	1201	TCTGTCCACACCGCTAGGTAGACAGCGTGCCTGAGACTCTTCGAGAGCCACACTGCTCG	1260
Db	1201	TCTGTCCACACCGCGTAAGTAGACAGCGCTGCGTCAAGACTCTTCGAGAGCCACACTGCTCG	1260
QY	1261	GCAGGATTATTCGATTCACGGATCAGTCAGACGAGGGTCCAGGAGCGGCGCTGACGGCACG	1320
Db	1261	GCAGGATTATTCGATTCACGGATCAGTCAGACGAGGGTCCAGGAGAGCGGCGCTGACGGCACG	1320
QY	1321	GGTGGCGGGG	1330
Db	1321	GGTGGCGGGG	1330

```

1      RESULT      4
2      US-08-868-288A-2
3      ; Sequence 2, Application US/08668288A
4      ; Patent No. 5922567
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Au-Yang, Janice
7      ; APPLICANT: Lal, Preeti
8      ; APPLICANT: Bandman, Olga
9      ; TITLE OF INVENTION: TWO NEW HUMAN DNAJ-LIKE PROTEINS
10     ; NUMBER OF SEQUENCES: 7
11     ; CORRESPONDENCE ADDRESS:
12     ; ADDRESSEE: Incyte Pharmaceuticals, Inc.
13     ; STREET: 3174 Porter Drive
14     ; CITY: Palo Alto
15     ; STATE: CA
16     ; COUNTRY: USA
17     ; ZIP: 94304
18     ; COMPUTER READABLE FORM:
19     ; MEDIUM TYPE: Diskette
20     ; COMPUTER: IBM Compatible
21     ; OPERATING SYSTEM: DOS
22     ; SOFTWARE: FastSeq for Windows Version 2.0
23     ; CURRENT APPLICATION DATA:
24     ; APPLICATION NUMBER: US/08/868,288A
25     ; FILING DATE: June 3, 1997
26     ; PRIORITY APPLICATION DATA:
27     ; APPLICATION NUMBER:
28     ; FILING DATE:
29     ; ATTORNEY/AGENT INFORMATION:
30     ; NAME: Billings, Lucy J.
31     ; REGISTRATION NUMBER: 36,749
32     ; REFERENCE/DOCKET NUMBER: PF-0309 US
33     ; TELECOMMUNICATION INFORMATION:
34     ; TELEPHONE: 415-855-0555
35     ; TELEFAX: 415-845-4166
36     ; INFORMATION FOR SEQ ID NO: 2:
37     ; SEQUENCE CHARACTERISTICS:
38     ; LENGTH: 1376 base pairs
39     ; TYPE: nucleic acid
40     ; STRANDEDNESS: single
41     ; TOPOLOGY: linear
42     ; IMMEDIATE SOURCE:
43     ; LIBRARY: SYNORAB01
44     ; CLONE: 136466
45     ; US-08-868-288A-2
46
47     Query Match      6.9%; Score 91.4; DB 4; Length 1376;
48     Best Local Similarity 66.5%; Pred. No. 5,5e-18;
49     Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;
50
51     QY      163 GATTCTATGAAAGTCTGAGCGCTGCAAGACATGCTCACCCGAGAGATTTTAAAAAGCA 222
52     DB      268 GATTTCTATAGATCTTGGGGGTGCTCGAAGTGCCTTAATAAAGATTTTAAAAAGGCC 327
53     QY      223 TATCGAAGACTGGCACTGAAGTGGCATCCAGATAAAAAATCCTGAGAAATAAAGAGAAAGCA 282

```

```

RESULT      5
US-09-235-373-2
; Sequence 2, Application US/09235373
; Patent No. 6001598
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lu, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNAL-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: .DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/235,373
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/866,288
; FILING DATE: June 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-235-373-2

Query Match      6.9%; Score 91.4; DB 5; Length 1376;
Best Local Similarity 66.5%; Pred. NO.5.5e-18;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps

OY      163 GATTACTATGAAGTCTTACGAGCGTCAGACATCCCTCACCCGAGAGATATTAAAGGCA 222
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      268 GATTTCCTAAGATCTTGCGGGGTCCCTCGAAGTCCTCTTTAAAGATATTAAAAAGCC 327

OY      223 TATGGGAATCGCAGCTGAGTGCATCCAGATATAAAATCTCTGAGATTAAAGAAAGCA 262
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      328 TATGGAAGAACTAGGCCCTGACCTTCATCCCGAGCCGGAACCTG--ATGATCCCAAGCC 384
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

OY      283 GAGAGAAATTCAGGCAAGTAGTGGGAGGCAATGCAATGAGTGTGTGGATGCTTAAGAAAGCG 342
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT      6
US-09-388-993-2
; Sequence 2, Application US/09388993
; Patent No. 6043222
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; TITLE OF INVENTION: TWO NEW HUMAN DNA-LIKE PROTEINS
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388,993
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/868,288
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0309 US
; TELECOMMUNICATION INFORMATION:
; TELEFAX: 415-855-4166
; TELEPHONE: 415-845-4166
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1376 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: SYNORAB01
; CLONE: 136466
US-09-388-993-2

Query Match      6.9%; Score 91.4; DB 5; Length 1376;
Best Local Similarity 66.5%; Pred. No. 5.5e-18;
Matches 147; Conservative 0; Mismatches 71; Indels 3; Gaps 1;

163 GATTTCATGAAGTCTTAGGCGTCGACAGACATGCGCTCACCCGAGAGATTATTAAGGCA 222
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 268 GATTCCTCTTAACATCTTGGGGGTGCCTCGAAGTGCCTCTATTAAGAAATTAAAAAGCC 327
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

223 TATCGAAACCTGGCACTGAAGGCGATCCAGATTAATAAAATCCGAGAAATAAAGACGCA 282
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 328 TATAGAAACTAGCCCTGACGCTTATCCGACGGAAACCTG---ATGATCCACAAGCC 384
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

283 GAGAGAAATTCACAGCAAGTAGCGGAGCGATATGAAGTGTCTCGATCTTAAGAAACGG 342
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 385 CAGGAGAAATTCAGAGATCTGGGTGCTCTTATGAGGTTCTGACAGATGATGAGAAACGG 444
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

343 GACATCTATGACAATATGCGCAAAAGAAAGATTAAATGGCG 383
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 445 AAACAGTACGATCTTATGCTGACAGAGAGATTAAAAAGATGG 485
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```
RESULT 7
US-08-486-955A-6
; Sequence 6, Application US/08486955A
; Patent No. 5747299
; GENERAL INFORMATION:
; APPLICANT: FATHMAN, Garrison
; APPLICANT: BLOOM, Debra
; TITLE OF INVENTION: Anergy Genes
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: CA
; COUNTRY: US
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,955A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20015
; REFERENCE/DOCKET NUMBER: A59741-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-781-1989
; TELEFAX: 415-398-3249
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 672 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-486-955A-6

Query Match 5.7%; Score 75.6; DB 2; Length 672;
Best Local Similarity 58.5%; Pred. No. 2.2e-13;
Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;

QY 143 CAATCTGTAACATGCTGATTAATGATTAAGTTCTAGCGTCGACAGACATGCTCAC 202
DB 23 CACTCTCTACTCTCGGGGAATCGTTATACCATGTTCTTGACCTGGACAMGAATGCAACCT 82
QY 203 CCGAGATATTAAAGGATATCGGAACTGACACTGAAGTGCATCCAGATTAATAATC 262
DB 83 CAGATGACATTAAGGATCTTATCGGAAGCTGCCCTGAAGTATCAACCTGACAAAGAAC 142
QY 263 CTAGAGATTAAGAGAGACAGAGAGAAATTCAGAGAGTACGAGGAGCATATGAATGC 322
DB 143 CTGATTAACCGAGGCTGCAGAG---CAAGTTTAAGAGAGATTAACAACGACAGCATCT 199
QY 323 TGTGCGATGCTAGAAAGGAGCATATGACAAATATGCGCAAGAGATTAAATGTTG 382
DB 200 TGCACACACGCCACGAAAGAAACATTATGACAAAGTATGCTCGTGGGCTCTATGTGG 259
QY 383 GNGGNGNGGTGG 395
DB 260 CGGAGCAGTTTGG 272
```

```
RESULT 8
US-08-974-546-2
; Sequence 2, Application US/08974546
; Patent No. 5945287
```

```
;; GENERAL INFORMATION:
;; APPLICANT: Hillman, Jennifer L.
;; APPLICANT: Lal, Preeti
;; APPLICANT: Corley, Neil C.
;; APPLICANT: Shah, Puri
;; TITLE OF INVENTION: TWO HUMAN HEAT SHOCK PROTEIN HOMOLOGS
;; NUMBER OF SEQUENCES: 6
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Incyte Pharmaceuticals, Inc.
;; STREET: 3174 Porter Dr.
;; CITY: Palo Alto
;; STATE: CA
;; COUNTRY: USA
;; ZIP: 94304
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: Fastpro for Windows Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/974,546
;; FILING DATE: Filed Herewith
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0428
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 650-855-0555
;; TELEFAX: 650-845-4166
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2349 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; IMMEDIATE SOURCE:
;; LIBRARY: BRATITUT21
;; CLONE: 2525691
; US-08-974-546-2
```

```
Query Match 5.5%; Score 73.4; DB 4; Length 2349;
Best Local Similarity 60.8%; Pred. No. 2.2e-12;
Matches 138; Conservative 0; Mismatches 83; Indels 6; Gaps 1;

QY 163 GATTACTATGAAGTTCTAGCGTCGACAGACATGCTCACCCGAGATATTAAAGGCA 222
DB 111 GATTATTACAAGATTCTTGGAATCCCATCGGGGCCAAGAGATGACATCAAGAAACC 170
QY 223 TATGGAAACTGGCACTGAAGTGCATCCAGATTAATAATCCTGAGATTAAGAAAGCA 282
DB 171 TACCGAAGATGCGCTTGGAATGCCACCCAGACAAGAT-----AAGAACCCAACT 224
QY 283 GAGAGAAATTCAGCAAGTACGAGGAGCATATGAAGTCTGTGCGATGCTTAAGAAAGG 342
DB 225 GAGGAGAACTTAAGAGAGATTGCAAGAGGCTATGATGTGCTAAGTGAACCCCAAGAAAG 284
QY 343 GACATCTATGACAAATATGCGAAAGAGAGATTAAATGTTGNGNGG 389
DB 285 GGCCTGTATGACCAAGTATGGGAGGAGGAGGCTGGAAGACCGCGGTGG 331
```

```
RESULT 9
US-08-472-534-4
; Sequence 4, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard R
; APPLICANT: Martin, Denis
```

```
? TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
? TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
? NUMBER OF SEQUENCES: 6
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: Fish & Neave
? STREET: 1251 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: United States of America
? ZIP: 10020
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/472,534
? FILING DATE:
? CLASSIFICATION: 424
? ATTORNEY/AGENT INFORMATION:
? NAME: Haley Jr, James F
? REGISTRATION NUMBER: 27,794
? REFERENCE/DOCKET NUMBER: Biovac-2
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212-596-9000
? TELEFAX: 212-596-9090
? TELEX: 14-8367
? INFORMATION FOR SEQ ID NO: 4:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 4320 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: linear
? MOLECULE TYPE: DNA (genomic)
? HYPOTHEICAL: NO
? ANTI-SENSE: NO
? ORIGINAL SOURCE:
? ORGANISM: Streptococcus pneumoniae
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 3265..4320
? OTHER INFORMATION: /product= "NH2-terminal portion of
? OTHER INFORMATION: DNA J"
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 682..2502
? OTHER INFORMATION: /product= "Heat-Shock Protein 72"
? FEATURE:
? NAME/KEY: mat_peptide
? LOCATION: 682..2502
?
? US-08-472-534-4

Query Match 5.1%; Score 68.2; DB 4; Length 4320;
Best Local Similarity 56.6%; Pred. No.1.2e-10;
Matches 146; Conservative 0; Mismatches 106; Indels 6; Gaps 1;
```

```
QY 394 GGAGTCATTTTGACAGT 411
DB 3502 GCTGTGTTTCGGCGGT 3519

RESULT 10
US-08-879-260-3
? Sequence 3, Application US/08879260
? Patent No. 5935851
? GENERAL INFORMATION:
? APPLICANT: Murthy, Anita E.
? APPLICANT: Gusella, James F.
? TITLE OF INVENTION: TPR-containing Genes
? NUMBER OF SEQUENCES: 11
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
? STREET: 1100 New York Ave, N.W., Suite 600
? CITY: Washington
? STATE: D.C.
? COUNTRY: U.S.A.
? ZIP: 20005
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patentin Release #1.0, Version #1.30
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/879,260
? FILING DATE: 19JUN1997
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER: 60/020,204
? FILING DATE: 20JUN1996
? ATTORNEY/AGENT INFORMATION:
? NAME: Ludwig, Steven R.
? REGISTRATION NUMBER: 36,203
? REFERENCE/DOCKET NUMBER: 0609,4260001/JNG/SRL
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 202-371-2540
? TELEFAX: 202-371-2540
? INFORMATION FOR SEQ ID NO: 3:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 1756 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 27..1478
?
? US-08-879-260-3

Query Match 5.0%; Score 66.4; DB 4; Length 1756;
Best Local Similarity 61.2%; Pred. No.2.4e-10;
Matches 131; Conservative 0; Mismatches 71; Indels 12; Gaps 1;
```

RESULT 11
US-08-221-816B-1
; Sequence 1, Application US/08221816B
; Patent No. 5738985
; GENERAL INFORMATION:
; APPLICANT: Miles, Vincent J.
; APPLICANT: Mathews, Michael B.
; APPLICANT: Katze, Michael G.
; APPLICANT: Witherell, Gary
; APPLICANT: Watson, Julia C.
; TITLE OF INVENTION: METHOD FOR SELECTIVE INACTIVATION
; TITLE OF INVENTION: OF VIRAL REPLICATION
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennile & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036/2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/221,816B
; FILING DATE: 01-Apr-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Coruzzi, Laura A
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7960-030
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1515 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; US-08-221-816B-1

Query Match 4.5%; Score 60; DB 2; Length 1515;
Best Local Similarity 59.0%; Pred. No. 1.9e-08;
Matches 124; Conservative 0; Mismatches 80; Indels 6; Gaps 1;

QY 150 GTAACAATGATGATTTACTATGAGTTCAGCGTCGACAGACATGCTCACCAGGGA 209
DB 1167 GTACACAGACGAGATTATATACAAATCTTGGAGTAAAGAAATGCCAAAGAGCA 1226
QY 210 TATTAAAAAGCATATCGGAACTGGCACTGAAGTGGCATCCAGATAA-----AATCC 263
DB 1227 AATCATTTAAAGCATACCGAAATTAACACTGACGTGACGCCACCCAGACACTTCCAGAACGA 1286
QY 264 TGAGATTAAGAAGAGACGAGAGAAATTCAGAAAGTAGCGGAGCATATGAGAGTGGT 323
DB 1287 AGAAGAAAAAGAAAAAGCTGAGAGAAAGTTTCATTTGACATAGCAGTGTAAAGAGTCTT 1346
QY 324 GTCCGATGCTAAGAAAGGAGCATCTATGA 353
DB 1347 CTCGATCCAGAAATGAGAGAAAGATTGA 1376

RESULT 12
US-08-143-219-26
; Sequence 26, Application US/08143219
; Patent No. 5670330

; GENERAL INFORMATION:
; APPLICANT: Sosenberg, Nahum
; APPLICANT: Katze, Michael G.
; APPLICANT: Roy, Sophie
; APPLICANT: Koromilas, Antonis E.
; APPLICANT: Barber, Glen N.
; TITLE OF INVENTION: TUMOR-CELL ASSAY METHOD AND KIT
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 611 West Sixth Street
; CITY: Los Angeles
; STATE: CA
; COUNTRY: USA
; ZIP: 90017
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 MB
; COMPUTER: IBM compatible
; OPERATING SYSTEM: PC-DOS (Version 5.0)
; SOFTWARE: Wordperfect (Version 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/143,219
; FILING DATE: October 25, 1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below:
; APPLICATION NUMBER: 08/141,244
; FILING DATE: October 22, 1993
; APPLICATION NUMBER: 07/953,681
; FILING DATE: September 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Douglas E. Olson
; REGISTRATION NUMBER: 22,798
; REFERENCE/DOCKET NUMBER: 204/139
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 26:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1687 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: BOVINE P58 GENE, FIGURE 10A
; US-08-143-219-26

Query Match 4.5%; Score 60; DB 1; Length 1687;
Best Local Similarity 59.0%; Pred. No. 2.1e-08;
Matches 124; Conservative 0; Mismatches 80; Indels 6; Gaps 1;

QY 150 GTAACAATGATGATTTACTATGAGTTCAGCGTCGACAGACATGCTCACCAGGGA 209
DB 1282 GTACACAGACGAGATTATATACAAATCTTGGAGTAAAGAAATGCCAAAGAGCA 1341
QY 210 TATTAAAAAGCATATCGGAACTGGCACTGAAGTGGCATCCAGATAA-----AATCC 263
DB 1342 AATCATTTAAAGCATACCGAAATTAACACTGACGTGACGCCACCCAGACACTTCCAGAACGA 1401
QY 264 TGAGATTAAGAAGAGACGAGAGAAATTCAGAAAGTAGCGGAGCATATGAGAGTGGT 323
DB 1402 AGAAGAAAAAGAAAAAGCTGAGAGAAAGTTTCATTTGACATAGCAGTGTAAAGAGTCTT 1461
QY 324 GTCCGATGCTAAGAAAGGAGCATCTATGA 353
DB 1462 CTCGATCCAGAAATGAGAGAAAGATTGA 1491

	Query Match	4.2%; Score 55.8; DB 1;	Length 7216;
	Best Local Similarity	3.5%; Pred. NO. 9.5e-07;	
Matches	9; Conservative	162; Mismatches	87; Indels
			Gaps 0;
Qy	141 AACATCTCGTAAACATGCTGGATTACTGTGAAGTTTCAGCGGTGCAGACATGCC TC	200	
Db	1447 AAGAATTGTTGACRR RR	1387	
Qy	201 ACCGGAGATATTTAAAAAGCATATCGAATACTGCACATGAAGTGCCATCCAGATAAAA	260	
Db	1387 RRR RR	1328	
Qy	261 TCTGTGAGATAAAGAACAGACAGAGAAAATTCAGAACAATGACGGAGGCATATGAAGT	320	
Db	1327 RRR RR	1268	
Qy	321 GCTGTGCGATGCTAAGAAACGGACATCTATGCAATTTGCCAAGAAGATTAAATGG	380	
Db	1267 RRR RR	1208	
Qy	381 TGNGGNGAGGTGAAG	398	
Db	1207 RRRRRRRRRRRRRRRR	1190	

```

Query Match Similarity      4.2%   Score 55.2; DB 4; Length 1700;
Best Local Similarity       57.9%;   Pred. No. 5.9e-07;
Matches 124; Conservative  0; Mismatches 78; Indels 12; Gaps 1

QY 154 AACATGCTGGATTTACTATGAAGTTCGTAGCGGTGCACAGACATGCTCACCCGAGATATT 213
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 983 AAGAGAGAAATTAATTACTACAAGATCCTGGAGATGGAGCAAMAATAAGCCCTTGAGAGACGAGATC 1042

QY 214 AAAAAGCATATCGGAACACTGGCACTGMAAGTGGCACTCCAGATRAAAA-----T 261
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1043 AAGAAAGCTTACCGBAACGGCGCTTGATGCAACCACTCCAGATCGGCACAGTGGGCCACT 1102

QY 262 CCTGGAATTAAGAAGAACAGACAGAGAAAAATTCAAGACATAGCGAGGCATATGAACTG 321
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1103 GCCGAAGTTCAAGAACGAGAGAGAGANAAGTTTAAGAAGATGGAGAGAGCCTTTACATC 1162

QY 322 CTGTGCGATGCTAAGAAAACGGGACATCTATGACA 355
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1163 CTCCTGTATCCCACAGAAAAAGACTGTTATGACA 1196

RESULT 15
US-08-686-417-2
; Sequence 2, Application US/08686417
; Patent No. 5850018
; GENERAL INFORMATION:

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 30, 2000, 14:10:37 ; Search time 70.1 Seconds
(without alignments)
4746.868 Million cell updates/sec

Title: US-09-501-714-4

Perfect score: 1330
Sequence: 1 CGNAGGAGGAGNAGGAGAG.....TGACGCGACGCGTGGCGGG 1330

Scoring table: IDENTITY_NTC
Gapop 10.0 , Gapext 1.0

Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : N_Geneseq_36.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1307.2	98.3	1330	X06100	Human DnaJ-like pr
2	91.4	6.9	1376	X06099	Human DnaJ-like pr
3	90.6	6.8	1204	X14194	H. pylori GHPO 885
4	84.2	6.3	6336	V174530	Staphylococcus aur
5	81.6	6.1	2717	X13034	Enterococcus faeca
6	79.6	6.0	1191	Q57429	DNA J like protein
7	75.6	5.7	672	T11789	Mouse cysteine str
8	75.6	5.7	672	T11789	Human cysteine str
9	69.4	5.2	110000	X20248_03	Continuation (4 of
10	68.2	5.1	1726	T73388	Homo sapiens fetal
11	68.2	5.1	4320	T73388	DNA encoding HSP72
12	68.2	5.1	7174	V52322	Streptococcus pneu
13	67.2	5.1	331	V86438	EST clone AM610. N
14	66	5.0	110000	T58840_0	Mycoplasma genital
15	65.6	4.9	336	X40691	Human secreted pro
16	65.6	4.9	28295	X20507	Polynucleotide seq
17	63	4.7	450	Q03633	Mycoplasma hyopneu
18	62.4	4.7	491	V02136	Human secreted pro
19	62.4	4.6	110000	T88058	Partial cDNA clone
20	61.2	4.6	110000	X20248_02	Continuation (3 of
21	60	4.5	1515	Q77721	p58 cDNA clone. SC
22	60	4.5	1515	V22052	Bovine p58 DNA. SC
23	60	4.5	1687	V01072	Bovine p58 gene. S
24	56.8	4.3	1701	X39655	Renal cancer assoc
25	56	4.2	459	V87477	EST clone BN379. N
26	55.2	4.2	1700	V11858	Mus musculus Tub I
27	54	4.1	3748	T84330	Maze DnaJ-related
28	54	4.1	3748	V72905	Maze DnaJ clone D
29	53.8	4.0	110000	T58840_2	Continuation (3 of
30	43.4	3.3	482	V86780	EST clone BF211. N
31	41.8	3.1	1516	V59688	Human secreted pro
32	41.4	3.1	828	X30633	H. pylori cytoplas
33	41.4	3.1	837	X30634	H. pylori cytoplas
34	39	2.9	4444	X20635	Polynucleotide seq

35	38.2	2.9	990	X14165	H. pylori GHPO 542
36	37.8	2.8	114955	X53491	Human adenosine Al
37	37.4	2.8	2500	V84584	Human secreted pro
38	35.8	2.7	1686	Q87587	DNA encoding Leuco
39	35.4	2.7	110000	V21209_13	Continuation (14 o
40	35.2	2.6	2811	T45660	Rhodospseudomonas s
41	35.2	2.6	3399	T05868	Chicken leucocytoz
42	35	2.6	35515	X20252	Borrelia burgdorfer
43	34.8	2.6	1956	T67161	Plasmodium falcipa
44	34.8	2.6	6225	X55273	Human enzyme-relat
45	34.4	2.6	13987	T80415	Hydrid srncg/tylg O

ALIGNMENTS

RESULT	ID	Location/Qualifiers
1	X06100	
AC	X06100: standard; DNA; 1330 BP.	
DT	09-APR-1999 (first entry)	
DE	Human DnaJ-like protein, HSP72 encoding DNA.	
KW	DnaJ-like protein; heat shock protein; HSP71; HSP72; cancer; leukaemia; immune disorder; inflammation; tissue damage; diabetes; wound healing; chromosome mapping; ss.	
KW	Homo sapiens.	
OS	Homo sapiens.	
FH	Key	
FT	157. .1149	
FT	Location/Qualifiers	
FT	/*tag= a	
FT	/product= "DnaJ-like protein, HSP72"	
FT	/transl_except= (pos:871. .873; aa:Xaa)	
FT	/transl_except= (pos:970. .972; aa:Xaa)	
FT	/transl_except= (pos:1024. .1026; aa:Xaa)	
FT	/transl_except= (pos:1027. .1029; aa:Xaa)	
FT	/transl_except= (pos:1033. .1035; aa:Xaa)	
FT	/transl_except= (pos:1048. .1050; aa:Xaa)	
FT	/transl_except= (pos:1057. .1059; aa:Xaa)	
FT	/transl_except= (pos:1060. .1062; aa:Xaa)	
FT	/transl_except= (pos:1069. .1071; aa:Xaa)	
FT	/transl_except= (pos:1081. .1083; aa:Xaa)	
FT	/note= "Xaa = unknown"	
W0985509-A2.		
10-DEC-1998.		
02-JUN-1998; U11182.		
03-JUN-1997; US-868288.		
(INCY-) INCYTE PHARM INC.		
Au-Young J, Bandman O, Lal P;		
WPI: 99-070259/06.		
P-PSDB; W94066.		
New nucleic acid encoding human DnaJ-like proteins - for diagnosis, treatment and prevention of cancer, immune disorders and inflammation		
PS Claim 27: Fig 3A-D; 73pp: English.		
This DNA encodes a human DnaJ-like protein, HSP72. The invention provides two human DnaJ-like proteins which are heat shock proteins J1 and J2 (HSPJ1 and HSPJ2) and nucleic acid sequences encoding the proteins. Host cells containing a vector comprising the nucleic acids are used for the production of recombinant proteins. Recombinant HSPJ1 and HSPJ2 are used to raise Ab, therapeutically and to screen for specific binding agents.		
Antagonists are used to treat or prevent a wide variety of solid cancers, leukaemia and lymphoma; immune disorders (typical of many disclosed are acquired immune deficiency syndrome, allergy, asthma, Crohn's disease, diabetes mellitus, multiple sclerosis, rheumatoid arthritis, microbial and parasitic infections) and inflammation. Agonists may be used to treat or prevent tissue damage, e.g. in cases of ankylosing spondylitis, heart attack, ischaemia, UV exposure, insulin-dependent diabetes and for wound healing, and may be expressed from a vector containing the nucleic acids.		
Fragments of the nucleic acids are used as primers and probes for detecting and quantifying the HSPJ-encoding nucleic acid in usual hybridisation and/or amplification assays, therapeutically as antisense, triplex-forming or ribozyme molecules, and for chromosome mapping.		
Sequence 1330 BP; 336 A; 318 C; 381 G; 271 T;		

Sequence 6336 BP; 1733 A; 1218 C; 839 G; 2364 T;

Query Match 6.3%; Score 84.2; DB 1; Length 6336;
Best Local Similarity 58.6%; Pred. No. 6.2e-14;
Matches 184; Conservative 0; Mismatches 121; Indels 9; Gaps 2;

QY 163 GATTCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 222
DB 3160 GATTCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 3101
QY 223 TATCGGAAGTGGCAGCTGAGTGCATCCAGATATAAATCTGAGATAAAGAGAGCA 282
DB 3100 TATCGGAAGTGGCAGCTGAGTGCATCCAGATATAAATCTGAGATAAAGAGAGCA 3047
QY 283 GAGAGAAATTCAGACAGTACCGGAGCATATGAGTGTGGATGCTAAGAAAGCG 342
DB 3046 GATGAAAGTTCAGAAATTCAGACAGTACCGGAGCATATGAGTGTGGATGCTAAGAAAGCG 2987
QY 343 GACATCTATGACAAATATGCAAGAGATTAATGCTGNGGNGGTGGA---AGT 399
DB 2986 GCAAGCTATGATCAATTTGGACATGATGCTCCACAGATTTTGGGCAAGAGATTCAT 2927
QY 400 CATTTGACAGTCCATTTGAAATTTGGCTTCACATTCGTAACCAAGATGATGCTCAGG 459
DB 2926 GGCCTGCTGACTTTGGCGGTTTAAAGTTCGTCGTCGCGCGGTTTGAAGATATCTTAGC 2867
QY 460 GAATTTTGTGGTGG 473
DB 2866 TCTTTCTTCGCTGG 2853

RESULT 5

ID X13034/C
X13034 standard; DNA; 2717 BP.
AC X13034;
DT 19-MAR-1999 (first entry)
DE Enterococcus faecalis genome contig SEQ ID NO:97.
KW Enterococcus faecalis; contig; detection; Enterococcal infection;
KM vaccine; attenuation; computer readable medium; ds.
OS Enterococcus faecalis.
PN M09850555-A2.
PD 12-NOV-1998.
PF 04-MAY-1998; U08985.
PR 14-NOV-1997; US-066009.
PR 06-MAY-1997; US-044031.
PR 16-MAY-1997; US-046655.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Barash SC, Dillon PJ, Kunsch CA;
DR WPI; 99-045171/04.
PT New isolated Enterococcus faecalis polynucleotides and polypeptides
PT - used to develop products for the detection of Enterococcus and for
PT use in vaccines for prevention or attenuation of Enterococcus
PT infection.
PS Claim 1: Page 632-634; 2084pp; English.
CC A computer readable medium has been developed which has recorded on it
CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.
CC X12938 to X13919 represent these nucleotide sequences which are primary
CC nucleotide sequences, also known as contigs. The computer-based system
CC can identify fragments of the Enterococcus faecalis genome with
CC commercial importance. The products can be used to detect the presence
CC of Enterococcus faecalis in samples. They can also be used for
CC diagnosing Enterococcal infection in an animal and monitoring
CC progression of disease, and for identifying agents which can be used to
CC modulate the growth or pathogenicity of Enterococcus faecalis, or
CC another related organism, in vivo or in vitro. In particular the
CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences
CC can be used in vaccines to prevent or attenuate an Enterococcal
CC infection.
SQ Sequence 2717 BP; 700 A; 580 C; 462 G; 971 T;

Query Match 6.1%; Score 81.6; DB 1; Length 2717;

Best Local Similarity 62.2%; Pred. No. 2.2e-13;
Matches 145; Conservative 0; Mismatches 82; Indels 6; Gaps 1;

QY 163 GATTCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 222
DB 321 GATTCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 262
QY 223 TATCGGAAGTGGCAGCTGAGTGCATCCAGATATAAATCTGAGATAAAGAGAGCA 282
DB 261 TACCGCAAACTTCCAAAAGATACATCCAGATTA-----TTAAACAAGAGCCGATGCG 208
QY 283 GAGAGAAATTCAGACAGTACCGGAGCATATGAGTGTGGATGCTAAGAAAGCG 342
DB 207 GAGAGAAATTCAGACAGTACCGGAGCATATGAGTGTGGATGCTAAGAAAGCG 148
QY 343 GACATCTATGACAAATATGCAAGAGATTAATGCTGNGGNGGTGGA---AGT 399
DB 147 GCGGCTATGACCAATATGCGCATGAGAGAACTGATCTTAATTCGCTGCGG 95

RESULT 6

ID O57429
O57429 standard; cDNA to mRNA; 1191 BP.
AC O57429;
DT 19-OCT-1994 (first entry)
DE DNA J like protein.
KW Human cDNA; library; enzyme; protein; ss.
OS Homo sapiens.
PN M09403599-A.
PD 17-FEB-1994.
PF 04-AUG-1993; J01095.
PR 04-AUG-1992; JP-208077.
PR 13-NOV-1992; JP-327619.
PR 26-FEB-1993; JP-061431.
PA (SAGA) SAGAMI CHEM RES CENTRE.
PI Iwahori A, Kato S, Kato T, Kim N, Oh S, Sekine S;
DR WPI; 94-065688/08.
DR P-PSDB; R46090.
PT cDNA of human origin and proteins coded by it - which may be
PT expressed by in vivo or in vitro translation using sense RNA or
PT antisense DNA corresponding to the cDNA.
PS Claim 1: Page 44-47; 167pp; Japanese.
CC mRNA expressed in human fibrosarcoma cell line HT-1080 was
CC isolated and used to construct a cDNA library using vector
CC pKAL. Clone HP0067 encoding DNA J-like protein
CC was isolated.
SQ Sequence 1191 BP; 397 A; 207 C; 309 G; 278 T;

Query Match 6.0%; Score 79.6; DB 1; Length 1191;
Best Local Similarity 62.1%; Pred. No. 5.2e-13;
Matches 146; Conservative 0; Mismatches 77; Indels 12; Gaps 1;

QY 165 TTAATCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 224
DB 18 TTAATCTATGAGTCTCTAGCGGTGAGACATGCTCCACCGGATATATAAAGGCA 77
QY 225 TCGGAAGTGGCAGCTGAGTGCATCCAGATATAAATCTGAGATAAAGAGAGCA 284
DB 78 TCGGAAGTGGCAGCTGAGTGCATCCAGATATAAATCTGAGATAAAGAGAGCA----- 130
QY 285 GAGAGAAATTCAGACAGTACCGGAGCATATGAGTGTGGATGCTAAGAAAGCG 344
DB 130 ----AAGTTTAAAGATTTCTCAAGCTTACAGATTTCTCTGTGTAAGAAAGAGCA 185
QY 345 CATCTATGACAAATATGCAAGAGATTAATGCTGNGGNGGTGGAAGT 399
DB 186 ATTATATGACAAAGAGAGAGAGCAATTAAGAGGTGAGAGAGGTGGCGGT 240

RESULT 7

Query Match 7
T11788 standard; cDNA; 672 BP.

AC T11788;
 DT 07-APR-1996 (first entry)
 DE Mouse cysteine string protein gene.
 KW Mouse; cysteine string protein; anergy; T-lymphocyte;
 KW differential display; cDNA; reverse transcription; probe;
 KW polymerase chain reaction; cloning; gel electrophoresis;
 KW antibody; diagnostic; immunoassay; autoimmune disease;
 KW organ transplantation; vaccine; drug screening; ss.
 OS Mus musculus.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bloom D, Fathman G;
 DR WPI: 96-068884/07.
 PT Identifying genes associated with T-cell anergy, and the proteins
 PT encoded by them - useful to elucidate the anergic state, and
 PS identify agents associated with induction of anergy
 PS Disclosure; Fig 2; 31pp; English.
 CC The sequence encodes a mouse cysteine string protein isolated from
 CC mouse T-cell clones 11.3.7 and 12.2 by a new procedure for
 CC identification of genes associated with anergy. The protein is
 CC expressed in resting T-cells but not in anergic T-cells. To
 CC isolate proteins of this type, differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.
 SQ Sequence 672 BP; 171 A; 173 C; 181 G; 147 T;

Query Match 5.7%; Score 75.6; DB 1; Length 672;
 Best Local Similarity 58.5%; Pred. No. 5.3e-12;
 Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;
 QY 143 CAATCTCGTAAACATGCGTGTGATTCTATGAAGTTCTAGCGGTGCGAGACATGCTCAC 202
 DB 23 CACTCTCTACTTCGCGGAATCGTTATACATGTTCTTGACTGACAAAGATGCAACT 82
 QY 203 CCGAGATATTAAAAAGGCATATCGGAACCTGCACTGAAGTGGCATCCAGTAAAAATC 262
 DB 83 CAGATGACATTAAAAAGCTTATCGGAAGCTGCGCCCTGAAGTATCACCTGCAAGAACC 142
 QY 263 CTGAATTAAGAAAGACAGAGAAAATTCAGCAAGTAGCGGAGCATATGAAGTGC 322
 DB 143 CTGATTAACCCAGAGGCTGCAGAAA--CAAGTTTAAGAGATTAAACACCGCATCT 199
 QY 323 TGTGGATGCTTAAGAACGGAGCATCTATGACAAATATGCAAAAGAGATTAAATGGTG 382
 DB 200 TGACAGAGCCGACGAAAGAAACATTATATGACAAAGTATGCGTGGGGGCTATATGG 259
 QY 383 GNGNGGNGGTGG 395
 DB 260 CGAGACAGTTTGG 272

RESULT 8
 T11789
 ID T11789 standard; cDNA; 672 BP.
 AC T11789;
 DT 07-APR-1996 (first entry)
 DE Human cysteine string protein gene.
 KW Human; cysteine string protein; anergy; T-lymphocyte;

KW differential display; cDNA; reverse transcription; probe;
 KW polymerase chain reaction; cloning; gel electrophoresis;
 KW antibody; diagnostic; immunoassay; autoimmune disease;
 KW organ transplantation; vaccine; drug screening; ss.
 OS Homo sapiens.
 PN W09600300-A1.
 PD 04-JAN-1996.
 PF 22-JUN-1995; U07958.
 PR 23-JUN-1994; US-265100.
 PA (STRD) UNIV LELAND STANFORD JUNIOR.
 PI Bloom D, Fathman G;
 DR WPI: 96-068884/07.
 PT Identifying genes associated with T-cell anergy, and the proteins
 PT encoded by them - useful to elucidate the anergic state, and
 PS identify agents associated with induction of anergy
 PS Disclosure; Page 19-20; 31pp; English.
 CC The sequence encodes a human cysteine string protein, the absence
 CC of which is associated with anergy in T-cells. Genes with at least
 CC 80% homology to this sequence (e.g. the mouse sequence T11788) may
 CC be used as probes for detection of the anergic state, and may be
 CC isolated by a new procedure. Differential display is first
 CC performed by generating cDNA from polyA+ mRNA with an oligo-dT
 CC primer with 2 variable anchor nucleotides at the 3'-end, e.g.
 CC T11786, using reverse-transcriptase. The resulting cDNA
 CC populations are amplified by the polymerase chain reaction,
 CC using the same oligo-dT primer and a 5'-primer which includes
 CC common 4-5 nucleotide sequence combinations, e.g. T11785. After
 CC gel electrophoresis of the products, bands unique to anergised or
 CC non-anergised cells are excised, eluted, amplified, cloned and
 CC sequenced. Genes and proteins (and corresponding antibodies)
 CC arising from this procedure may be used e.g. in monitoring
 CC autoimmune disease, organ transplantation, vaccine
 CC effectiveness, and in drug screening.
 SQ Sequence 672 BP; 171 A; 173 C; 181 G; 147 T;

Query Match 5.7%; Score 75.6; DB 1; Length 672;
 Best Local Similarity 58.5%; Pred. No. 5.3e-12;
 Matches 148; Conservative 0; Mismatches 102; Indels 3; Gaps 1;
 QY 143 CAATCTCGTAAACATGCGTGTGATTCTATGAAGTTCTAGCGGTGCGAGACATGCTCAC 202
 DB 23 CACTCTCTACTTCGCGGAATCGTTATACATGTTCTTGACTGACAAAGATGCAACT 82
 QY 203 CCGAGATATTAAAAAGGCATATCGGAACCTGCACTGAAGTGGCATCCAGTAAAAATC 262
 DB 83 CAGATGACATTAAAAAGCTTATCGGAAGCTGCGCCCTGAAGTATCACCTGCAAGAACC 142
 QY 263 CTGAATTAAGAAAGACAGAGAAAATTCAGCAAGTAGCGGAGCATATGAAGTGC 322
 DB 143 CTGATTAACCCAGAGGCTGCAGAAA--CAAGTTTAAGAGATTAAACACCGCATCT 199
 QY 323 TGTGGATGCTTAAGAACGGAGCATCTATGACAAATATGCAAAAGAGATTAAATGGTG 382
 DB 200 TGACAGAGCCGACGAAAGAAACATTATATGACAAAGTATGCGTGGGGGCTATATGG 259
 QY 383 GNGNGGNGGTGG 395
 DB 260 CGAGACAGTTTGG 272

RESULT 9
 X20248_03
 Continuation (4 of 10) of X20248 from base 300001 (Borrelia burgdorferi polynucleotide
 WP Sequence split into 10 fragments LOCUS X20248 Accession X20248
 WP Fragment Name Begin End
 WP X20248_00 110000
 WP X20248_01 100001 210000
 WP X20248_02 200001 310000
 WP X20248_03 300001 410000
 WP X20248_04 400001 510000
 WP X20248_05 500001 610000
 WP X20248_06 600001 710000

[illegible]

QY	DB	108	TACATGTATCTTACGTAGGTCGCAAAATTCGGCATCAGAGGCCCAATCATAAAGGCCCTT	167
QY	226	CGGAACGTGCACGAACTGGCATTCACGATTAATAATCTTGAAATTAAGAAGACGACAG	285	
DB	168	CACAGTGTGGCAAGAGTACACCTCGACAAAAAT-----AAGAGCCCGATGCTGAA	221	
QY	286	AGAAATTTCAAGCAGTAGCGGAGCATATGATGAATGCTGTGAGATGCTTAAGAACGGGAC	345	
DB	222	GCAAAATTTCAAGAGATTTGCGAAGCATATCAAACTCTAGATGCTTAATAGCGAA	281	
QY	346	ATCATATGACAAATATGCGAAGAAGATTTAATGCTGGGNGGNGGTGGAAT	399	
DB	282	GAGTATGATACACTTGGACACAGTGTCTTTACTGTGTAAGGACAAAGAGT	335	
RESULT	11			
T73388				
ID	T73388	standard; DNA; 4320 BP.		
AC	T73388;			
DT	01-OCT-1997	(first entry)		
DE	DNA encoding HSP72 and N-terminus of DnaJ protein.			
KW	Streptococcus pneumoniae; S. pyogenes; S. agalactiae; genome; HSP72;			
KW	open reading frame; heat shock protein 72; DnaJ; DnaK; chimeric;			
KW	E. coli; Lactococcus lactis; antibody; vaccine; infection; human; ds.			
OS	Streptococcus pneumoniae.			
FS	Location/Qualifiers			
FT	key			
FT	cds	682..2505		
FT		/tag= a		
FT		/gene= "DnaK"		
FT		/product= heat shock protein 72		
FT		3265..4320		
FT	cds	/tag= b		
FT		/product= N-terminus of DnaJ protein		
PN	WO640928-A1.			
PN	19-DEC-1996.			
PF	17-MAY-1996; CA03322.			
PR	07-JUN-1995; US-472534.			
PR	04-AUG-1995; US-001805.			
PI	(IABF-) IAF BIOVAC INC.			
PI	Brodeur B, Hamel J, Martin D, Rioux C;			
DR	WPI; 97-052328/05.			
DR	P-PSDB: W22357-22358.			
PT	Streptococcal heat shock proteins and corresponding DNA sequences -			
PT	used in the production of a vaccine to treating and preventing			
PT	strain-specific Streptococcal infection			
PS	Claim 28; Page 90-94; 156pp; English.			
CC	This nucleotide sequence, isolated from a Streptococcus pneumoniae			
CC	HindIII-partially digested genomic DNA library contains 2 open reading			
CC	frames. The first encodes the heat shock protein 72 (HSP72) and the			
CC	second is a partial sequence encoding the N-terminus of the DnaJ protein.			
CC	The sequence was isolated using a fragment of the chimeric gene (W73392			
CC	corresponding to the sequence encoding the C-terminal 169 amino acids of			
CC	HSP72. The HSP72 gene encodes a protein of calculated molecular weight			
CC	of 64755 daltons (as compared to 72 kD by SDS-PAGE), with a pI of 4.35.			
CC	The protein has 58% and 85% identity to the E. coli and Lactococcus			
CC	lactis HSP72 proteins respectively. The HSP72 protein and its fragment,			
CC	or antibodies specific to HSP72, are used in pharmaceutical compositions,			
CC	pref. a vaccine, for treating or preventing infection by S. pneumoniae or			
CC	related bacteria in humans, e.g. S.pyogenes or S. agalactiae.			
SQ	Sequence 4320 BP; 1402 A; 845 C; 971 G; 1102 T;			
Query Match	5.1%;	Score 68.2;	DB 1;	Length 4320;
Best Local Similarity	56.6%;	Pred. No. 1.8e-09;		
Matches 146;	Conservative 0;	Mismatches 106;	Indels 6;	Gaps 1.
QY	154	AACATGCTGATATCTTACGATTAAGTTCTAGCGCTGCAGACAGATGCTTCACCCGAGATATT	213	
DB	3268	AACATGCTGATATCTTATGATGATGCTGTGGGGGATATCCAAAAACGCTTGGCAGACGAATTC	3327	
QY	214	AAAAAGCATATGCGAAACTGGCACTGAAGTGGCATCCAGATTAATAATCTTGAGATATA	273	
DB	3328	AAAAAGGCTTATCGTACCTTCCAAAAAATATTCACCCAGATATACACAGAGGCT---	3385	


```
RESULT 14
T58840_0
WP Sequence split into 6 fragments LOCUS T58840 Accession T58840
WP Fragment Name Begin End
WP T58840_0 1 110000
WP T58840_1 100001 210000
WP T58840_2 200001 310000
WP T58840_3 300001 410000
WP T58840_4 400001 510000
WP T58840_5 500001 580073
ID T58840 standard; DNA; 580073 BP.
AC T58840;
DE 27-MAR-1997 (first entry)
DE Mycoplasma genitalium genome.
KM megabase shotgun sequencing method; open reading frame; ORF; ss.
OS Mycoplasma genitalium.
FH Key Location/Qualifiers
FT cds
FT /*tag= a
FT /label= MG006
FT /note= "Previously identified as MORF-20076, the
FT encoded protein shows 27.59 percentage
FT identity to thymidylate kinase (CDC8)
FT from Saccharomyces cerevisiae"
FT 11252..12040
FT /*tag= b
FT /label= MG009
FT /note= "Previously identified as MORF-20078, the
FT encoded protein shows 35.43 percentage
FT identity to the Bacillus subtilis hypothetical
FT protein covered in accession number
FT GB:D26185.102"
FT 12069..12725
FT /*tag= c
FT /label= MG010
FT /note= "Previously identified as MORF-20079, the
FT encoded protein shows 25.73 percentage
FT identity to DNA primase (dnaE) from
FT Clostridium acetobutylicum"
FT complement (13570..14247)
FT /*tag= d
FT /label= MG012
FT /note= "Previously identified as MORF-20080, the
FT encoded protein shows 31.50 percentage
FT identity to the ribosomal protein S6
FT modification protein (rimK) from Escherichia
FT coli"
FT complement (14396..15217)
FT /*tag= e
FT /label= MG013
FT /note= "Previously identified as MORF-19823, MORF-20080
FT and MORF-20081, the encoded protein shows 33.04
FT percentage identity to 5,10-methylene-tetra-
FT hydrofolate dehydrogenase (tdld) from E. coli"
FT 17474..19243
FT /*tag= f
FT /label= MG015
FT /note= "Previously identified as MORF-20084, the
FT encoded protein shows 32.23 percentage
FT identity to transport ATP-binding protein
FT (msbA) from E. coli"
FT 26478..27344
FT /*tag= g
FT /label= MG023
FT /note= "Previously identified as MORF-20092, the
FT encoded protein shows 45.96 percentage
FT identity to fructose-bisphosphate aldolase
FT (tsr) from B. subtilis"
FT 27345..28448
FT /*tag= h
FT /label= MG024
```

```
FT /note= "Previously identified as MORF-19826 and
FT MORF-20093, the encoded protein shows 46.84
FT percentage identity to GTP-binding protein
FT from E. coli"
FT 36987..38978
FT /*tag= i
FT /label= MG032
FT /note= "Previously identified as MORF-20099, the
FT encoded protein shows 26.82 percentage
FT identity to ATP-dependent nuclease (addA)
FT from B. subtilis"
FT 35242..35904
FT /*tag= j
FT /label= MG033
FT /note= "Previously identified as MORF-20100, the
FT encoded protein shows 35.90 percentage
FT identity to glycerol uptake facilitator
FT (g1pF) from B. subtilis"
FT complement (39873..40514)
FT /*tag= k
FT /label= MG034
FT /note= "Previously identified as MORF-20101, the
FT encoded protein shows 48.13 percentage
FT identity to thymidylate kinase (tdk)
FT from B. subtilis"
FT 40543..41787
FT /*tag= l
FT /label= MG035
FT /note= "Previously identified as MORF-20102, the
FT encoded protein shows 30.71 percentage
FT identity to histidyl-tRNA synthetase (h1ss)
FT from Mycobacterium leprae"
FT complement (44751..46277)
FT /*tag= m
FT /label= MG038
FT /note= "Previously identified as MORF-20105, the
FT encoded protein shows 46.83 percentage
FT identity to glycerol kinase (g1pK)
FT from E. coli"
FT complement (46268..47422)
FT /*tag= n
FT /label= MG039
FT /note= "Previously identified as MORF-19831 and
FT MORF-20106, the encoded protein shows 43.20
FT percentage identity to glycerol-3-phosphate
FT dehydrogenase (gdr2) from S. cerevisiae"
FT 49377..49643
FT /*tag= o
FT /label= MG041
FT /note= "The encoded protein shows 48.86 percentage
FT identity to phosphohistidinoprotein-hexose
FT phosphotransferase (ptsH) from Mycoplasma
FT capricolum"
FT 50060..51520
FT /*tag= p
FT /label= MG042
FT /note= "Previously identified as MORF-19832 and
FT MORF-20108, the encoded protein shows 41.92
FT percentage identity to spermidine/
FT putrescine transport ATP-binding protein
FT (potA) from E. coli"
FT 51525..52382
FT /*tag= q
FT /label= MG043
FT /note= "Previously identified as MORF-20110, the
FT encoded protein shows 26.51 percentage
FT identity to spermidine/putrescine transport
FT system permease protein (potB) from E. coli"
FT 52366..53220
FT /*tag= r
FT /label= MG044
FT /note= "Previously identified as MORF-20111, the
FT encoded protein shows 29.45 percentage
```


Query Match	Best Local Similarity	Matches 142; Conservative	5.0%; Score 66; DB 1; Length 110000;	Pred. No. 4.3e-08;	Mismatches 93;	Indels 12;	Gaps 11;
FT	identity to spermidine/putrescine transport						
FT	system permease protein C (potc) from E. coli"						
FT	54658. .55605						
FT	cds						
FT	/*tag= s						
FT	/label= MG046						
FT	/note= "Previously identified as MORF-20112, the						
FT	encoded protein shows 36.60 percentage						
FT	identity to stialoglycoprotease (gcp)						
FT	from Pasteurella haemolytica"						
FT	complement (56970. .58310)						
FT	cds						
FT	/*tag= t						
FT	/label= MG048						
FT	/note= "Previously identified as MORF-19834,						
FT	MORF-20114 and MORF-20115, the encoded protein						
FT	shows 43.02 percentage identity to signal						
FT	recognition particle protein (ifh) from B.						
FT	subtilis"						
FT	58117. .59079						
FT	cds						
FT	/*tag= u						
FT	/label= MG049						
FT	/note= "Previously identified as MORF-20114 and						
FT	MORF-20115, the encoded protein shows 44.78						
FT	percentage identity to purine-nucleoside						
FT	phosphorylase (decd) from E. coli"						
FT	59083. .59754						
FT	cds						
FT	/*tag= v						
FT	/label= MG050						
FT	/note= "Previously identified as MORF-20117, the						
FT	encoded protein shows 83.03 percentage						
FT	identity to deoxyribose-phosphate aldolase						
FT	(deoc) from Mycoplasma pneumoniae"						
FT	complement (64898. .65731)						
FT	cds						
FT	/*tag= w						
FT	/label= MG056						
FT	/note= "Previously identified as MORF-20122, the						
FT	encoded protein shows 30.25 percent						
FT	identity to the protein disclosed in						
FT	GB:D26185..99 from B. subtilis"						
FT	complement (65713. .66249)						
FT	cds						
FT	/*tag= x						
FT	/label= MG057						
FT	/note= "Previously identified as MORF-20123, the						
FT	encoded protein shows 38.90 percentage						
FT	identity to the protein disclosed in						
FT	GB:D26185..104 from B. subtilis"						
FT	81047. .82597						
FT	cds						
FT	/*tag= y						
FT	/label= MG067						
FT	/note= "Previously identified as MORF-19845, the						
FT	encoded protein shows 28.84 percentage						
FT	identity to glutamic acid specific protease						
FT	(Spase) from Staphylococcus aureus"						
FT	91065. .91919						
FT	cds						
FT	/*tag= z						
FT	/label= MG070						
FT	/note= "Previously identified as MORF-20136, the						
FT	encoded protein shows 34.8 percentage						

QY	330	TGCTAAGAAAGCGGACACCTATGACAAATATGCGCAAGAGATTAATATGCGGNGGNGG	389
Db	22585	TGAAGAAAAAGCTAAGCTTTATGACCAAGTTTGTCATGCAAGGGTTAAATGCTTCTGCTTT	22644
QY	390	NGGTGCA 396	
Db	22645	TCATGAA 22651	
RESULT	15		
ID	X40691		
AC	X40691	standard; cDNA; 336 BP.	
DT	18-JUN-1999	(first entry)	
DE	Human secreted protein 5' EST SEQ ID NO: 291.		
KW	Human; secreted protein; EST; expressed sequence tag; diagnosis;		
KW	forensic; gene therapy; chromosome mapping; signal peptide; prostate;		
KW	upstream regulatory sequence; cytokine activity; cell proliferation;		
KW	differentiation; haematopoiesis regulation; tissue growth regulation;		
KW	reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;		
OS	thrombolytic; anti-inflammatory; tumour inhibition; ds.		
PN	WO906550-A2.		
PD	11-FEB-1999.		
PF	31-JUL-1998; IB1232.		
PR	01-AUG-1997; US-905144.		
PA	(GIST) GENSET.		
PI	Duclet A, Dumas Milne Edwards J, Lacroix B;		
DR	WPI; 99-153780/13.		
DR	P-PSDB; Y11969.		
PT	New isolated prostate-derived nucleic acids - used to develop		
PT	products which may have cytokine, immune regulatory, haematopoiesis		
PT	regulating, anti-inflammatory or tumour inhibition activity		
PS	Claim 1; Page 455-456; 675pp; English.		
CC	X40438 to X40715 represent 5' expressed sequence tags (ESTs) for human		
CC	secreted proteins expressed in prostate, and encode the proteins given		
CC	Y11716 to Y11993 respectively. The proteins given represent the signal		
CC	peptide and an N-terminal fragment of a secreted protein. The nucleic		
CC	acid sequences can be used for producing secreted human gene products.		
CC	They can also be used to develop products for diagnosis and therapy. The		
CC	proteins obtained may have cytokine activity, cell proliferation and		
CC	differentiation activity, haematopoiesis regulating activity, tissue		
CC	growth regulating activity, reproductive hormone regulating activity,		
CC	chemotactic/chemokinetic activity, hemostatic and thrombolytic activity,		
CC	receptor/ligand activity, anti-inflammatory activity, tumour inhibition		
CC	activity or other activities. The products can be used in forensic, gene		
CC	therapy and chromosome mapping procedures. The sequences can also be used		
CC	for obtaining corresponding promoter sequences. The nucleic acids		
CC	encoding the signal peptides can be used for directing extracellular		
CC	secretion of a polypeptide or the insertion of a polypeptide into a		
CC	membrane, or importing a polypeptide into a cell.		
CC	Sequence 336 BP; 92 A; 77 C; 105 G; 60 T;		
QY	164	ATTACTATGAACTTACGGCTGCGAGACATCCTCACCGAGATATTAAAAAGCAT	223
Db	53	ACTACTATGACATCTCTGGCGCTGTAAGCCACGAGCGCTC -CCCGAGAGATCAAGAAGGCTT	111
QY	224	ATCGGAAATGCGACTGCAAGTGGCATCCGATATAAAAATCTCTGATATAAAGAAAGCAG	283
Db	112	ATCGGAGCTGGCGCTCAATGACACCCGCAAGAAACCCGGATGAGGCGAG-----	165
QY	284	AGAGAAATTCACAGCAAGTACGAGGACATATGAAAGTCTGTCGATGCTAAGAAACGGG	343
Db	165	-----AAGTTTAACTCATATCCACAGGCAATATGAAGTGTCTTCAGATCCAAAGAAAGG	219
QY	344	ACATCATGACAAATATATGCAAGAAAGATTTAAATGTCGNGGNGGNGG 392	

Db 220 ATGTTTATGACCAAGCGGAGAGCGCAATBVAAGAAGAGGCTCAGG 268

Search completed: June 30, 2000, 14:12:45
Job time: 8232 sec

Biochem. J. 284, 469-476, 1992
 A:Title: Human homologues of the bacterial heat-shock protein DnaJ are preferentially ex
 A:Reference number: S23508; MID:92287055
 A:Accession: S23508
 A:Status: Preliminary
 A:Molecule type: mRNA
 A:Residues: 1351 <CH>
 A:Cross-references: EMBL:X63368; NID:g32468; PIDN:CAA44968.1; PID:g32469
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:3-69/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 37.0%; Score 635.5; DB 2; Length 351;
 Best Local Similarity 54.4%; Pred. No. 7.8e-44;
 Matches 142; Conservative 30; Mismatches 48; Indels 41; Gaps 9;

QY 1 MVDYEVGVORHASPEDIKAYRKALKMHPDKNPENKEAEKRFQVAEAYVLSDAK 60
 Db 1 MASTYELDVPKASADIKAKRKALKMHPDKNPENKEAEKRFQVAEAYVLSDAK 60
 QY 61 KRDYDKYKKEGLNGGGGGSHFDSPP---EFGEFTFRNPDPVREFPGGDRPESFDFED 117
 Db 61 KREIYDRYEGELGLTGTSGRBAAGSGGPGFTFTFRSPREVEFRFGSGDDPFA-ELFDD 119
 QY 118 --PPEDFGRRRGRGSRSGTGSFSAFSGFPSSFGSGSFDTGTFSSGLHGILTSF 175
 Db 120 LGPFSEL--ONRG---SRHSGPFFTFSSSF-----GH---SDF 150
 QY 176 SSTSFSGS-GMGNPKISTSTKMWNGRKITTKRIVNGOERVEVEDGOLSLTINGVD 234
 Db 151 SSSSFSSPGAGAFRSYSTTTFYQGRITTKRLMNGOERVEVEDGOLSLTINGVD 210
 QY 235 DDALXERMRGQNVLPAPQA 255
 Db 211 DLARGLELSREQ-----GPS 226

RESULT 3
 JN0912
 heat-shock protein hsp40 - human
 C:Species: Homo sapiens (man)
 C:Date: 14-Jul-1994 #sequence_revision 14-Jul-1994 #text_change 20-Aug-1999
 C:Accession: JN0912; PNM0688; PH0843
 R:Ohtsuka, K.
 Biochem. Biophys. Res. Commun. 197, 235-240, 1993
 A:Title: Cloning of a cDNA for heat-shock protein hsp40, a human homologue of bacterial
 A:Reference number: JN0912; MID:94071949
 A:Accession: JN0912
 A:Molecule type: mRNA
 A:Residues: 1-340 <OHT1>
 A:Cross-references: DDBJ:D49547; DDBJ:DI749; NID:g710654; PIDN:BA08495.1; PID:d1009106
 A:Accession: PNM0688
 A:Molecule type: protein
 A:Residues: 3-49 <OHT2>
 A:Experimental source: Placenta
 R:Hattori, H.; Liu, Y.C.; Tohnai, I.; Ueda, M.; Kaneda, T.; Kobayashi, T.; Tanabe, K.; O
 Cell Struct. Funct. 17, 77-86, 1992
 A:Title: Intracellular localization and partial amino acid sequence of a stress-inducibl
 A:Reference number: PH0843; MID:92266438
 A:Accession: PH0843
 A:Molecule type: protein
 A:Residues: 'X', '3-46', 'Y', '48-49 <HAT>
 A:Experimental source: HeLa cell
 C:Genetics:
 A:Gene: GDB:HSPL1
 A:Cross-references: GDB:3929189
 A:Map position: 19p13.2-19p13.2
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 C:Keywords: heat shock; stress-induced protein
 F:4-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 21.2%; Score 364.5; DB 2; Length 340;

Best Local Similarity 32.9%; Pred. No. 4.8e-22;
 Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DVEYEVGVORHASPEDIKAYRKALKMHPDKNPENKEAEKRFQVAEAYVLSDAK 62
 Db 4 DYTDTGLARGASDEETKAKYRKALKMHPDKNPENKEAEKRFQVAEAYVLSDAK 61
 QY 63 DLYDKYKKEGL-----NGGGGGSHFDSPPREFGTFR-NPDYFREFGGDRPESFDFED 116
 Db 62 EIDFRYGEGLKSGPGSGGGGANGTS---FSYTFIDGPHAMFAEFGGR----- 110
 QY 117 DPEDFGRRRGRGSRSGTGSFSAFSGFPSSFGSGSFDTGTFSSGLHGILTSF 176
 Db 110 NPFDTFFGQRNGEGM-----DIDPFSGP-----MGMG---FT 142
 QY 177 STSFSGSGMGNPKISTSTKMWNGRKITTKRIVNGOERVEVE--DGLKSLTIN-- 231
 Db 143 NVNFGNS-----NSAQPAPAKKDDPVT-----HDLRVSLLEIYSGCTKKKISHKRL 190
 QY 231 ----GVADDALXERMRG-----QNVLPAPAGLRPPRPA5LLRHXPHCSKEE 280
 Db 191 NPDGKSRNMDKILITLVKKGWKEGRTITPKRGDDTSSNIPADIVFVLKDKRHNIFKRD 250
 QY 281 G 281
 Db 251 G 251

RESULT 4
 T09133
 heat shock protein homolog DNAJ - Trypanosoma brucei
 N:Alternate names: chaperrone
 C:Species: Trypanosoma brucei
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 20-Sep-1999
 C:Accession: T09133
 R:Bringaud, F.; Vedreone, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays,
 Mol. Biochem. Parasitol. 94, 249-264, 1998
 A:Title: Conserved organization of genes in trypanosomatids.
 A:Reference number: Z16580
 A:Accession: T09133
 A:Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A:Residues: 1-336 <BRI>
 A:Cross-references: EMBL:AF031926; NID:g3452211; PID:g3452212
 A:Experimental source: strain Anra1
 C:Genetics:
 A:Gene: dnaJ
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 C:Keywords: heat shock; molecular chaperrone; stress-induced protein

Query Match 21.2%; Score 363; DB 2; Length 336;
 Best Local Similarity 50.6%; Pred. No. 6.3e-22;
 Matches 84; Conservative 11; Mismatches 45; Indels 26; Gaps 5;

QY 2 VDVEYEVGVORHASPEDIKAYRKALKMHPDKNPENKEAEKRFQVAEAYVLSDAK 61
 Db 3 IDYKYVLGVSRDASPDIKKAYHQLKTHPDASGNREAEHLFEVAEAYVLSDEKK 62
 QY 62 RDYDKYKKEGLNGG-----GGGSHFDSPPREFGTFRNPD--DYFREFGGDRP 110
 Db 63 KRIYDSYGEGLKGVDPDSSGPGAGFGFSGGGGTYNFNSRDAFAKFEESFGSNDPF 122
 QY 111 SPDFEFDPDFEDFRGNRGRPGSRSGTGSFSAFSGFPSSFGSGSFDTGTFSSGLHGILTSF 176
 Db 123 AGG-----DMFGG--GP-----GLHRYFRGFGPGHGMGSGFS 153

RESULT 5
 T04618
 heat shock protein homolog F2009.160 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 20-Sep-1999

	Query Match	20.9%	Score 358;	DB 21;	Length 338;	
	Best Local Similarity	49.1%;	Pred. No.1.6e-21;			
	Matches	82;	Conservative	20;	Mismatches	39;
					Indels	26;
					Gaps	7.
QY	2 VDYEVVLGVORIASPEDIKKAYRKALKMHPDKNPENKEEAKRKFQVQAEVYLSDAKK	61				
Dd	3 IDIYVVLVGRRATSDSIKKATHQLALTKYHPPKCIGNRESERREKRVESEADYLDENK	62				
QY	62 RDIYDKYKEGGLNG-----GG--GGSHFDSPEFEG--FTFRMPD--DVFEFFGGGR	109				
Dd	63 KRIYDVYGEGLKGVPAGGEGSGAGGTGFHGAPFGGVRYTFSQGDAFIPIFSFGSSDP	122				
QY	110 FSFDFEDFDFFCGNRKRPGRSRSRGTSFSSAHSAGFSSFGSGRSS	156				
Dd	123 FAGG-----EEFGGG--GP-----GLHYRVRGCGPGGFTSGFS	155				

Query Match	19.38;	Score	332;	DB-2;	Length	178;			
Best Local Similarity	49.48;	Pred. No.	9.2e-20;						
Matches	76;	Conservative	16;	Mismatches	40;	Indels	22;	Gaps	6;

QY 3 DYEEVLGVQRHASPEDIKKAYRKLALKWHPDKNPENKKEAEERKKQVAEAYEVLSDAKKR 62

Db 4 DYICILGKIGKASDDEIKKAYRKQALKFHPDKN--KSPQAEKKEFEVAAEYVLSDPKKR 61
 QY 63 DIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDDPSPFDFEDPFD 122
 Db 62 ELYDQGEGLGAGAGCTGCGGTTRYP-HGDPHATFAAFYXGSNPFI-----F 111
 QY 123 FGNRRGPPGSRGSRG---TGSPFAF---SGFP 148
 Db 112 FGRRMG--GGRDSEMEIDGDPFSAFSGFSMNGYP 143

RESULT 9

H71526
 Probable heat shock protein 7 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
 C:Accession: H71526
 R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A:Reference number: A71570; MUID:99000809
 A:Accession: H71526
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-392 <ARN>
 A:Cross-references: GB:AE001307; GB:AE001273; NID:93328757; PIDN:AAC67936.1; PID:9332876
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: dnaJ
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:2-67/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 19.2%; Score 330; DB 2; Length 392;
 Best Local Similarity 37.3%; Pred. No. 3.4e-19;
 Matches 107; Conservative 27; Mismatches 89; Indels 64; Gaps 14;
 QY 2 VDYEVLGVORHASPEDIKKAYRKALKWHPDKNPKNEKEAEKRFKQVAEYVLSDAKK 61
 Db 1 MDYITLGVATATPAPEELKAYRKALKYHPDKNP-GDAEAEERFKEYSKAEYVLSDAQK 59
 QY 62 RDYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDDPSPFDFEDPFD-----F 114
 Db 60 RESIDRKGKGGPPFAGAGGFC-----GAGMGMEALRPFKMA---FGGDFGNGGGR 108
 QY 115 FEFDF---EDFFGNRRGPPGSRGSRG---TGSPFAFSGFPs--FGSGFSSFDGTFTS 163
 Db 109 FEGFLGGLGEAFGMRGSESERQASKKVHTLSFEAAKGVKEKLLVSGYKSCD----- 164
 QY 164 FGSILGHGGLSFSSTSG-----GSGM-----GNFKSISTSTKMV-NGRKITTIRIV 209
 Db 164 -ACGSGG---ANTAKGVVCDCKCKSGGVVOSRGFFSNASTCPCDSCGGRVITDPCSV 217
 QY 210 ENGGERVEVEDQLSLTINGVADDAALKEERMR---RGONVLP 252
 Db 218 CRGGGRK-----DKRSVHNITAGVDSGMRKLMKEGIGDAGONGAPRA 259

RESULT 10

F71623
 Protein with DnaJ domain PFB0090C - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 26-Aug-1999
 C:Accession: F71623
 R:Gardner, M.J.; Tettelin, H.; Garucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.; Perteau, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O. Science 282, 1156-1132, 1998
 A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A:Reference number: A71600; MUID:99021743
 A:Accession: F71623
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA

A:Residues: 1-354 <GAR>
 A:Cross-references: GB:AE001370; GB:AE001362; NID:93845089; PID:93845091; TIGR:PFB009
 A:Experimental source: clone 307
 C:Genetics:
 A:Gene: PFB0090C
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:21-89/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 19.1%; Score 327; DB 2; Length 354;
 Best Local Similarity 37.6%; Pred. No. 5.3e-19;
 Matches 85; Conservative 23; Mismatches 52; Indels 66; Gaps 8;

QY 3 DYVEVLGVORHASPEDIKKAYRKALKWHPDK--NPNKKEAEKRFKQVAEYVLSDAK 60
 Db 21 DYIILGVSRDCTIEDIKKAYRKALKWHPDKHLNAASKKEADNMKRSISEAEYVLSDE 80
 QY 61 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDDPSPFDFEDPFE 120
 Db 81 KRDIYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGGDDPSPFDFEDPFE 126
 QY 121 DFGNRRGPPGSRGSRGSGFSAFSGFPSSFGSFDGTFTSGSLGHGGLTSFSTGS- 180
 Db 126 -----YNSPSSPNNGNVLFEE-----GSL-FGGSSPFGGIMP 155
 QY 180 FGGSGMGNFKSISTSTK-----WYNGR-----KITTKR 207
 Db 156 RSGSGYITTSKFSFSSMDKVEEYVPLVTLLEDLYNGIQKILKLYTRKR 201

RESULT 11

S34630
 dnaj protein homolog - human
 C:Species: Homo sapiens (man)
 C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 20-Aug-1999
 C:Accession: S34630; S34631
 R:Chellatiah, A.; Davis, A.; Mohanakumar, T. Biochim. Biophys. Acta 1174, 111-113, 1993
 A:Title: Cloning of a unique human homologue of the Escherichia coli DnaJ heat shock A:Accession: S34630; MUID:93326629
 A:Reference number: S34631; MUID:93326630
 A:Molecule type: mRNA
 A:Residues: 1-397 <CHE>
 A:Cross-references: EMBL:L08069; NID:9306713; PIDN:AAC37517.1; PID:9306714
 A:Experimental source: umbilical vein endothelium
 R:Oh, S.; Iwahori, A.; Kato, S. Biochim. Biophys. Acta 1174, 114-116, 1993
 A:Title: Human cDNA encoding DnaJ protein homologue.
 A:Reference number: S34631; MUID:93326630
 A:Accession: S34631
 A:Molecule type: mRNA
 A:Residues: 1-273, 'H', 275-397 <JMA>
 A:Cross-references: EMBL:D13388; NID:9719587; PIDN:BA02656.1; PID:dl003160; PID:9219
 A:Experimental source: fibrosarcoma HT-1080
 C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology
 F:6-68/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 18.1%; Score 311; DB 2; Length 397;
 Best Local Similarity 32.8%; Pred. No. 1.2e-17;
 Matches 88; Conservative 30; Mismatches 72; Indels 78; Gaps 10;

QY 4 YVEVLGVORHASPEDIKKAYRKALKWHPDKNPKNEKEAEKRFKQVAEYVLSDAKKRD 63
 Db 7 YVDVLGVKNATQPEELKAYRKALKYHPDKNPNGE---KFKQISDAYEVLSDAKRE 62
 QY 64 IYDKYKKEGLNGGGGSHFDSPEFGFTFRNPDDVFEFGG-----RDPSPDFE 115
 Db 63 IYDKGGEQAIKRGAGG-----FGSPMDIFDMFGGGRMQRERGRKNVYHQL 111
 QY 116 EDPFEDFG-----NRGPPRS-----RSRGTGSPFAFSGFPSPFGS 152
 Db 112 SYTLEDLYNGATRKALKQKNVYCDCKSGRGGKAGVCECCPNCRGTMQI----- 161

```

Query Match          18.0%; Score 309; DB 2; Length 402;
Blast Local Similarity 229.4%; Pred. No. 1.7e-17;
Matches 91; Conservative 38; Mismatches 64; Indels 116; Gaps 11;

QY 4 YVEVLGVORHASPEDIKAYRKRLAKMHPDKNPENKEEAEKRFQVAAEYVLSDAKRKD 63
   ||:||||: ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 7 YVDVAGVPDASDNLKKAYRKMAKLFHPDKNPDAEQ---FKQISQAEYVLSDEKKRQ 62
   ||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:

QY 64 IYDKGKXELNCGGGGSGSHFDSPFEFGFTFRNPDDVFNFEFGGRPSPFDFEDFEDDF 123
   |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:
Db 63 IYDGGGEELQGGGGGGGGEFHNPF-----DVFDMFEG----- 95
   |||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 124 GNRRRPGRSGRSRSGTSPFSFAFGSPSPFGSGFSPFDTCGTSFGLHGGLT----SFSST 178
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 95 GGGRRGRGRRRYK-----PTVNHRLRYTLDTLYK-----GVTKKLKLSRAT 135
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 179 SFGSGMGNGFKSISTSTKMNCRKTTKRIYE----- 211
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 136 CKGCGGLCGNGESAKEGSCDRCRGRIKV-RVIMGPMVOMOSHCDSCNGEGSTPLEKDRC 194
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 211 ---NGQERYVEE-----DGLKSLTINQVAD-----DDALKEE 241
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 195 KKCNCRRKQVKEDEILVGLTPGMKQGE--KVFPEBKGDEVIQIEKPGDEYVVLDEVEHKK 252
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

QY 242 RMRRCQNVL 250
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

Db 253 FVRRKGDNLI 261
   ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||: ||:

RESULT 13
G72128
neat shock protein J - Chlamydia pneumoniae (strain CWI.029)
C:Species: Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 08-Oct-1999
C:Accession: G72128
R:Kaitman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A12000; MUID:99206606
A:Accession: G72128
A:Molecule type: DNA

```

```

Oy 2 VDYEVVLGVONHASEDIKKRYKRYLTKLMPHDKPEKKFEKREKFOVAAYELSDPAK 61
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MDYISILICISTASAEETKKRYKRLAYKHPDKP -GDAAAEKKFKKVSAYEYELSDP0K 59

Oy 62 RDIDYKYKEBGLGGGGGSHFSDPFEEGFT -FRNPDVYREFFG -GNDPFSPEFFED 117
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 60 RDSYDRFEGKSDPFAAGG - - - - - FGGAGGGMNEDLRTFPMGAFGEFGGSPFDG 110

Oy 118 PF--EDFFGNRRRPGGRSRSGTG - - - - - SFFAFSGFPS -FGSFSFDFGTFSGS 166
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 LFGGIGEFNGKSDPAGKROGASKRVHNLTFEBAHGVKEKELVSSYKCEET-CSOGA 169
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Oy 167 LGHGILTFSSSTSGGSGM - - - - - GNEKSISTSPKM -VNGRKILTKRIVNGOERYVEE 220
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 170 VNPQIKSCCEK - - - - - GSGQVYQSGNFGFMSMSTCEGCGEGRILITPDSCSRGGRVK - - - 225

Oy 221 DGLAKSLTINGVADDALKEBRM - - - - - RQONTPLA 252
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 225 --DKRSVHVH.IPAGVDSGMR.LKMGYDAGG.NGPS 258

```

RESULT 14
S42031
LDJ2 protein - leek
C:Species: Allium porrum (leek)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #test_change 20-Aug-1999
C:Accession: S42031
Ribessoule, J.-J.; Testet, E.; Cassagne, C.
submitted to the EMBL Data Library, February 1994
A:Reference number: S42031
A:Accession: S42031
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-418 <BES>
A:Cross-references: EMBL:X77632; NID:g454913; PIDN:CA54720.1; PID:g45430303
C:Superfamily: heat shock protein dnad; dnad amino-terminal homology <DN>
F113-74/Domain: dnad amino-terminal homology <DN>

[illegible]

Db 217 QOKKYLE-----VHEKGMQNGO--KITEPGEADE 244

RESULT 15

S26703

dnaJ protein homolog YDJ1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: dnaJ protein homolog MAS5; protein N2418; protein YNL064c; protein YN

C:Species: Saccharomyces cerevisiae

C:Date: 07-May-1993 #sequence-revision 07-May-1993 #text-change 20-Aug-1999

C:Accession: S26703; A39659; S58714; S62992; S17250

R:Atencio, D.P.; Yaffe, M.P.

Mol. Cell. Biol. 12, 283-291, 1992

A:Title: MAS5, a yeast homolog of dnaJ involved in mitochondrial protein import.

A:Reference number: S26703; MUID:92107179

A:Accession: S26703

A:Molecule type: DNA

A:Residues: 1-409 <ATP>

A:Cross-references: GB:S74758; NID:g241522; PIDN:AAB20771.1; PID:g241523

R:Caplan, A.J.; Douglas, M.G.

J. Cell Biol. 114, 609-621, 1991

A:Title: Characterization of YDJ1: a yeast homologue of the bacterial dnaJ protein.

A:Reference number: A39659; MUID:91332099

A:Accession: A39659

A:Molecule type: DNA

A:Residues: 1-409 <CAP>

A:Cross-references: GB:X56560; NID:g4810; PIDN:CAA39910.1; PID:g4811

R:Bergez, P.; Dolignon, F.; Crouzet, M.

Yeast 11, 967-974, 1995

A:Title: The sequence of a 44 420 bp fragment located on the left arm of chromosome XIV

A:Reference number: S58711; MUID:96021608

A:Accession: S58714

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-409 <BER>

A:Cross-references: EMBL:U12141; NID:g1314216; PIDN:AAA9647.1; PID:g994823

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1994

R:Bergez, P.; Dolignon, F.; Crouzet, M.

submitted to the Protein Sequence Database, April 1996

A:Reference number: S62975

A:Accession: S62992

A:Molecule type: DNA

A:Residues: 1-409 <BEW>

A:Cross-references: EMBL:Z71340; NID:g1301940; PIDN:CAA95937.1; PID:e239683; PID:g130194

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:YDJ1; MAS5

A:Cross-references: SGD:S0005008; MIPS:YNL064c

A:Map position: 14L

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

C:Keywords: heat shock; mitochondrion; stress-induced protein

F:6-70/Domain: dnaJ amino-terminal homology <DNM>

Query Match 17.4%; Score 298.5; DB 2; Length 409;

Best Local Similarity 29.7%; Pred. No. 1,2e-16;

Matches 95; Conservative 36; Mismatches 86; Indels 103; Gaps 14;

QY 4 YVEVLGVQRIASPEDIKAVRKALAKWHPDKNPENKEEAERKFOYAEAYEVLSDAKKRD 63

Db 7 FYDILGVPTATDYVEIKKAYKCALKHPDKNP--SEAAEKFKKASAAVEILSDPEKRD 64

QY 64 IYDKYGEKGLNGGGGSHDSPFEFGFTFRNDDVFRFFGGRDPSFFEDFEDFF 123

Db 65 IYDQFGEDGLSGAGAGGFGGFGFGG-----DDIFSQFFGA----- 102

QY 124 GNRGRPRG--SRSRGTGSFFSFGSFGSFGSFGSFGSL-----GHGGL 172

Db 102 GGAQRPRGPRGKDIKHEISA-----SLEELYKGRTRAKLAKQILCKECBERGCK 152

QY 173 TSF--STSFSGSGM-----GNFKSISTSTRMNGRKITTK 206

Db 153 KGAVKKCTSCNGGIGKIVTRQMGPMIORFQTECDVCHGTGDIIDPRCKSCSKKKEVNE 212

QY 207 R-IYVENGERVEYVE---EDGOLKSITINGVADDALAEERRRRQGNVLPAPAGLRPPKP 262

Db 213 RKILE-----VHEPGRMDGQ--RIVEKGEAD-----CAPDIVIPGD----- 247

QY 263 PRPASLRLHXPHCLSKERGE 282

Db 247 --VVEIYSERPHKSPKRDGD 264

Search completed: June 30, 2000, 09:58:20

Job time: 940 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 11:56:22 ; Search time 40.12 Seconds
(without alignments) 250.502 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYVEVLCVGRHASPEDIK.....EAEKGVEEEDVQROSLDRT 330

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 83857 seqs, 30454973 residues

Total number of hits satisfying chosen parameters: 83857

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SwissProt_38.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1230	71.7	241	1	HSJ2_HUMAN
2	1158.5	67.5	242	1	HSJ2_MOUSE
3	847.5	49.4	242	1	HSJ3_MOUSE
4	641.5	37.4	351	1	HS43_HUMAN
5	364.5	21.2	340	1	HS41_HUMAN
6	361	21.0	334	1	DNJ1_DROME
7	311	18.1	397	1	DNJ2_HUMAN
8	305	17.8	397	1	DNJ2_MOUSE
9	304	17.7	369	1	DNJ1_LEPIN
10	302.5	17.6	418	1	DNJ2_ALLPO
11	298.5	17.4	409	1	MAS5_YEAST
12	295	17.2	390	1	DNJ1_MYCPN
13	293	17.1	372	1	DNJ1_BACSU
14	287	16.7	413	1	DNJH_CUCSA
15	286.5	16.7	374	1	DNJ1_CIOAB
16	284	16.6	370	1	DNJ1_ERYRH
17	283.5	16.5	234	1	DNJ1_RHILE
18	281.5	16.4	375	1	DNJ1_BRIOV
19	281.5	16.4	389	1	DNJ1_MYGE
20	280.5	16.3	369	1	DNJ1_NITREU
21	280.5	16.3	417	1	DNJH_ABTNU
22	280	16.3	419	1	DNJH_AKATH
23	279.5	16.3	378	1	DNJ1_SALTY
24	278	16.2	367	1	DNJ1_COXBU
25	277	16.1	379	1	DNJ1_LACIA
26	276	16.1	352	1	P25294_saccharomyc
27	276	16.1	371	1	DNJ1_YEAST
28	274.5	16.0	380	1	DNJ1_BACST
29	271.5	15.8	379	1	DNJ1_LEGPN
30	271	15.8	375	1	DNJ1_ECOLI
31	270.5	15.8	389	1	DNJ1_METMA
32	268.5	15.6	332	1	DNJH_SYNY3
33	266.5	15.5	384	1	DNJ1_RHOCA
34	264.5	15.4	382	1	DNJ1_HAEIN

35	263.5	15.4	377	1	DNJ1_HAEIN	P48208 haemophilus
36	263	15.3	379	1	DNJ1_STAU	P45555 straphylococ
37	261.5	15.2	511	1	MDJ1_YEAST	P35191 saccharomyc
38	260.5	15.2	364	1	DNJ1_BORBU	P28616 borrelia bu
39	259.5	15.1	352	1	DNJ1_STRPN	P95830 streptococc
40	257.5	15.0	397	1	DNJ1_ALLPO	P03363 allium porr
41	256	14.9	404	1	SCJ1_YEAST	P25303 saccharomyc
42	254.5	14.8	355	1	YRVL_YEAST	P25303 saccharomyc
43	251.5	14.7	379	1	PSI_SCHPO	Q10005 caenorhabdi
44	246.5	14.4	370	1	DNJ1_RICPR	Q09912 schizosacch
45	245.5	14.3	375	1	DNJ1_ACTAC	G9zdy0 rickettsia P77866 actinobacil

ALIGNMENTS

RESULT	ID	Sequence	Standard	PRT	241 AA.
HSJ2_HUMAN	075190	O95806			
AC	15-FEB-2000	(Rel. 39, Created)			
DT	15-FEB-2000	(Rel. 39, Last sequence update)			
DT	15-FEB-2000	(Rel. 39, Last annotation update)			
DE	DNJ1 PROTEIN HOMOLOG 2 (HEAT SHOCK J2 PROTEIN) (HSJ-2) (MSJ-1).				
GN	HSJ2 OR MSJ1				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;				
CC	Eutheria; Primates; Catarrhini; Hominoidea; Homo.				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-TESTIS:				
RA	Saito T., Seki N.;				
RT	"A new member of human dnaj-related gene family."				
RL	Submitted (MAY-1998) to the EMBL/Genbank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Zhang W., Wan T., Yuan Z., Cao X.;				
RT	"HSJ2, a novel human homologue of the bacterial heat-shock protein				
RL	Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-TESTIS:				
RX	MEDLINE: 99115663.				
RA	Pel L.;				
RT	"Pituitary tumor-transforming gene protein associates with ribosomal				
RL	protein S10 and a novel human homologue of Dnaj in testicular cells."				
J.	Biol. Chem. 274:3151-3158(1999).				
CC	-1- SUBUNIT: INTERACTS WITH PTTG.				
CC	-1- SIMILARITY: CONTAINS A DNJ-LIKE DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: AB014888; BAA32209.1; -				
DR	EMBL: AF075601; AAD43194.1; -				
DR	EMBL: AF080569; AAD16010.1; -				
DR	HSSP: P25685; 1HDJ.				
DR	PFAM: PF00226; Dnaj. 1.				
DR	PROSITE: PS00636; DNJ1_1; 1.				
KW	Chaperone.				
FT	DOMAIN				
FT	CONFICT				
FT	SEQUENCE				
FT	241 AA; 26900 MW; CD5A49CAB66C2E16 CRC64;				
FT	LDNR -> WMAEKSLQRELSRTYKKE (IN REF. 3).				
FT	SVNGRKTTTKRIVNGGKRVVEEDGQLSLTINGKQLLR				
FT	Query Match				
FT	71.7%; Score 1230; DB 1; Length 241;				

Best Local Similarity 100.0%; Pred. No. 1.2e-87;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
DB 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
QY 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 120
DB 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 120
QY 121 DEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 180
DB 121 DEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 180
QY 181 GSGSGMGNFKSISTSTKMWNGRKITTKRIYENGOERVEVEEDGQLKSLTING 231
DB 181 GSGSGMGNFKSISTSTKMWNGRKITTKRIYENGOERVEVEEDGQLKSLTING 231

RESULT 2
HSJ2_MOUSE
ID HSJ2_MOUSE STANDARD; PRT; 242 AA.
AC 054946;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DNAB PROTEIN HOMOLOG 2 (HEAT SHOCK J2 PROTEIN) (HSJ-2) (MRJ).
GN HSJ2 OR MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthelia; Rodentia; Sciurognath; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-129; TISSUE-EMBRYONIC TROPHOBLAST;
RC Hunter P.J., Swanson B.J., Haendel M., Lyons G.E., Cross J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS A DNAB-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF035962; AAC16759.1; -
DR HSSP; P25685; 1HDJ.
DR PRAM; PF00226; DNAB_1; 1.
DR PROSITE; PS00636; DNAB_1; 1.
KW Chapterone.
FT DOMAIN 1 69 DNAB-LIKE.
SQ SEQUENCE 242 AA; 27012 MW; BAAC2D9EB26F0FA4 CRC64;

Query Match 67.5%; Score 1158.5; DB 1; Length 242;
Best Local Similarity 93.5%; Pred. No. 3.7e-82;
Matches 217; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
DB 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
QY 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 119
DB 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 120
QY 120 EDEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 179
DB 121 DEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 180
QY 180 FGSGSGMGNFKSISTSTKMWNGRKITTKRIYENGOERVEVEEDGQLKSLTING 231

DB 181 FGSGSGMGNFKSISTSTKMWNGRKITTKRIYENGOERVEVEEDGQLKSLTING 232

RESULT 3
HSJ3_MOUSE
ID HSJ3_MOUSE STANDARD; PRT; 242 AA.
AC 035723;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DNAB PROTEIN HOMOLOG 3 (HEAT SHOCK J3 PROTEIN) (HSJ-3) (MSJ-1).
GN HSJ3 OR MSJ1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Euthelia; Rodentia; Sciurognath; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA TISSUE-TESTIS;
RC MEDLINE; 98189155.
RX Berruti G., Borgonovo B., Martegani E.;
RT "MSJ-1, a new member of the DNAB family of proteins, is a male germ
RT cell-specific gene product."
RL Exp. Cell Res. 239:430-441(1998).
CC -1- TISSUE SPECIFICITY: TESTIS-SPECIFIC.
CC -1- SIMILARITY: CONTAINS A DNAB-LIKE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U95607; AAC13944.1; -
DR HSSP; P25685; 1HDJ.
DR MGD; MGI:1306822; HSJ3.
DR PRAM; PF00226; DNAB_1; 1.
DR PROSITE; PS00636; DNAB_1; 1.
KW Chapterone.
FT DOMAIN 1 69 DNAB-LIKE.
SQ SEQUENCE 242 AA; 26678 MW; 5F54F48A0094D84 CRC64;

Query Match 49.4%; Score 847.5; DB 1; Length 242;
Best Local Similarity 70.0%; Pred. No. 2.6e-58;
Matches 163; Conservative 30; Mismatches 37; Indels 3; Gaps 3;

QY 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
DB 1 MVDYEVLGVOHRAHPEDIKKAYRKALKWHPDKNPENKEERKQVAEYVLSDAK 60
QY 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 118
DB 61 KRDIDYKYEGLNGGGGGSHFDSPEFEGFTFRNPDDVFRFEGGRDPFSDFEDPFE 120
QY 119 DEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 178
DB 121 DEFGRRGPRGSRSGTSFSAFSGFSGSSFTGTFSGSLGHGGLTSSSTFS 180
QY 179 FGSGSGMGNFKSISTSTKMWNGRKITTKRIYENGOERVEVEEDGQLKSLTING 231
DB 180 SCGGGAAGMYKSVSTETINCKITTKRIYENGOERVEVEEDGQLKSLTING 232

RESULT 4
HS43_HUMAN
ID HS43_HUMAN STANDARD; PRT; 351 AA.
AC P25686;
DT 01-MAY-1992 (Rel. 22, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)

FT CONFLICT 183 183 M -> T (IN REF. 1).
 FT CONFLICT 320 320 V -> A (IN REF. 1).
 SQ SEQUENCE 340 AA; 38044 MW; 17545098B0C196DF CRC64;

Query Match
 Best Local Similarity 21.2%; Score 364.5; DB 1; Length 340;
 Matches 99; Conservative 46; Mismatches 81; Indels 75; Gaps 13;

QY 3 DYEVVLGVORHASPEDIKKAAYRKALAKWHPDKNPNKNEKEAEKFKOVAEAYEVLSDAKK 62
 DB 4 DYVYVLGLARASDEIKKAYRKALRHHPDKNK- -PGAEEKKEIAEAYEVLSDPKR 61
 QY 63 DIYDYKKEGL- - - - -NCGGGGSHFSDPFERGFTR- -NPDDVFERFGPDPFSDPFED 116
 DB 62 EIFFDYGEGLKSGSGSGGANGTS- - -FSYFFHDDPHAMFEFFGGR- - - - - 110
 QY 117 DPEDFPFRNRGRPSRSGRSGTSFFSATSFGSGSGSFSDTFTSGSLGHGGLSFS 176
 DB 110 NPFDFEFGQRNGEGM- - - - -DIDDPFGFP- - - - -MGMG- - - - -FT 142
 QY 177 STFGSGMGNFKSTSTKMWNGRKITTKRIVENGQERVEVE- - -DQGLSLTIN- - - 231
 DB 143 NVNNGRS- - - - -RSGQEPARKKQDPPTV- - - - -HDLRSLSELYSGCTKKMKISHRL 190
 QY 231 - - - - -GVADDALKEERRRG- - - - -QNVLPAPAGLRPPKPPASILRRHPCLSKEE 280
 DB 191 NPDGKSHNEDKILITEVKKMGKEGTKITPKEGDOTSNIPADIVFLKDKPHNIFKRD 250
 QY 281 G 281
 DB 251 G 251

RESULT 6
 DNJ1 DROME STANDARD; PRT; 334 AA.

AC 024133;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNJ1 PROTEIN HOMOLOG 1 (DROJ1).
 GN DNJ1-1 OR DROJ1.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;
 RN Ephydroidea; Drosophilidae; Drosophila.
 RP SEQUENCE FROM N.A.
 RA Lee J.Y., Palter K.B.
 RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- INDUCTION: BY HEAT SHOCK.
 CC -1- SIMILARITY: CONTAINS A DNJ1-LIKE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U34904; AAC23584.1; -
 DR HSSP: P25685; 1HDJ.
 DR FLYBASE: FBgn0015657; DnaJ-1.
 DR PROSITE: PS00636; DNJ1_1; 1.
 DR PROSITE: PS50076; DNJ2_1.
 DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF01556; DnaJ_C1.
 DR Pfam; Pf01556; DnaJ_C1.
 FT DOMAIN 1 70
 SQ SEQUENCE 334 AA; 37100 MW; 6056A235D6F2DC08 CRC64;

Query Match
 Best Local Similarity 21.0%; Score 361; DB 1; Length 334;
 Matches 99; Conservative 37; Mismatches 91; Indels 66; Gaps 9;

QY 3 DYEVVLGVORHASPEDIKKAAYRKALAKWHPDKNPNKNEKEAEKFKOVAEAYEVLSDAKK 62
 DB 4 DYVYVLGLARASDEIKKAYRKALRHHPDKNK- -SPAEEKKEIAEAYEVLSDPKR 61
 QY 63 DIYDYKKEGLNCGGGGSHFSDPFERGFTR- -NPDDVFERFGPDPFSDPFED 121
 DB 62 DIFDNYGEDGLKSGQPGDPGGGQCATYTFQHDPRATFQFFGSSDPF- - - - - 111
 QY 122 FFGNRKRPGRSGRSGTSGTSFFSATSFGSGSGSFSDTFTSGSLGHGGLSFSSTFG 181
 DB 111 - - - - -GAF- - - - -TGDDNFGSGGQGTNEIFMNIG 137
 QY 182 GSGMGNFKSTSTKMWNGRKITTKRIVENGQERVEVEEDQGLSLTINVA- -DDALX 239
 DB 138 GDDMFAPNAQAPSKKQDPPIEHDFVS- - -LEVDVGGCTKKKISMATGSGPYK 192
 QY 240 EERRRGQNVLPAPAGLR- - - - -PPKPPR- - - - -ASILRRHPCLSKEE 281
 DB 193 EEKYL-ITVKGPKWKGTKITPQEGDSAPNKPADIVFLIRKPHSLKREG 244

RESULT 7

DNJ2_HUMAN STANDARD; PRT; 397 AA.

AC P31689;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE DNJ1 PROTEIN HOMOLOG 2 (HSJ-2).
 GN HSJ2 OR DNJ2 OR HDJ2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 93326630.
 RA Oh S., Iwahori A., Kato S.;
 RT "Human cDNA encoding DnaJ protein homologue."
 RL Biochim. Biophys. Acta 1174:114-116(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE: 93326629.
 RA Chelliah A., Davis A., Mohanakumar T.;
 RT "Cloning of a unique human homologue of the Escherichia coli DnaJ
 RT heat shock protein."
 RL Biochim. Biophys. Acta 1174:111-113(1993).
 CC -1- SIMILARITY: CONTAINS A DNJ1-LIKE DOMAIN AND THE FOUR CXXCGXG
 CC MOTIFS FOUND IN DNJ1 PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D13388; BAA02656.1; -
 DR EMBL: L08069; AAC37517.1; -
 DR PIR: S34631; S34631.
 DR HSSP: P25685; 1HDJ.
 DR MIM: 602837; -
 DR PRINTS: PR00625; DNJ1PROTEIN.
 DR PROSITE: PS00636; DNJ1_1; 1.
 DR PROSITE: PS00637; DNJ1_CXXCGXG; 1.
 DR PROSITE: PS50076; DNJ2_1; 1.
 DR PFAM: PF00226; DnaJ_1.

CC EMBL: AF055664: AAC78597.1; -
DR EMBL: U53922: AAA9885.1; -.
DR HSSP: P25685: 1HDJ.
DR MGD: MGI:1270129: HSJ2.
DR PRINTS: PR00625: DNAJPROTEIN.
DR PROSITE: PS00636: DNAJ_1; 1.
DR PROSITE: PS00637: DNAJ_CXXCXXG; 1.
DR PROSITE: PS0076: DNAJ_2; 1.
DR PFAM: PF00226: DnaJ; 1.
DR PFAM: PF00684: DnaJ_CXXCXXG; 1.
DR PFAM: PF01556: DnaJ_C; 1.
KW Chapterone; Repeat.
FT DOMAIN 4 70 DNAJ-LIKE.
FT DOMAIN 75 96 GLY-RICH.
FT REPEAT 134 141 CXXCXXG MOTIF.
FT REPEAT 150 157 CXXCXXG MOTIF.
FT REPEAT 177 184 CXXCXXG MOTIF.
FT REPEAT 193 200 CXXCXXG MOTIF.
SQ SEQUENCE 397 AA; 44868 MW; 1783CE3D5C4CD558 CRC64;

Query Match 17.8%; Score 305; DB 1; Length 397;
Best Local Similarity 32.1%; Pred. No. 1,9e-16;
Matches 86; Conservative 32; Mismatches 72; Indels 78; Gaps 10;

QY 4 YEEVLGVGHASPEDIKKAYRKLLAKWHPDKNPENKEAEERFKQVAEYVLSDAKRD 63
DB 7 YYDVLGVAPNATQDEELKAYRKLLAKYHPDKNPNEGE---KFKQISQAEYVLADSKRRE 62
QY 64 IYDKYKRGKLGNGGGGGGSHDPSPEFGTFPRNPVDFEFKFG-----RDPSDF 115
DB 63 LYDGKGEGALIEGGAGG-----FSPMDIFDMFGGGGRMRERGRKNVHQL 111
QY 116 EDPEDFEG-----NRGRGRS-----BSRGNSEFSAFSGPSFGS 152
DB 112 SVTLEDLYNGATRIKLLALQKNVYCDKCEGRGKKGAVECCPMCRGTGMQI----- 161
QY 153 GFSSDFTGFTSFGSLGHGGLTFSSTSPFGSGMGNFKSIISTKMYNGRKIT-TRIVEN 211
DB 161 -----RIHQGLPGPMVOQIQISVCMECGHGHERISPKDRCKSCNGRKIVREKILE- 210
QY 212 GQERYVE---EDGQLKSLTINGVADD 236
DB 210 ----VHIDKGMKDGQ-KITFHEGDDQ 231

RESULT 9
DNAJ_LEPIN ID STANDARD; PRT; 369 AA.
AC P71443;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE DNAJ PROTEIN.
GN DNAJ.
OS Leptospira interrogans.
OC Bacteria; Spirochaetales; Leptospiaceae; Leptospira.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-WIJNBERG / SEROVAR COPENHAGEN1;
RX MEDLINE; 98382569.
RA BALLARD S.A., GO M., SEGERS R.P.A.M., ADLER B.;
RT "Molecular analysis of the dnaJ locus of Leptospira interrogans
serovar Copenhagen1";
RL Gene 216:21-29(1998).
RN [2]
RP SEQUENCE OF 1-35 FROM N.A.
RC STRAIN-SEROTYPE POMONA;
RA Gul B., Yellon D.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: INTERACTS WITH DNA, TO DISASSEMBLE A PROTEIN COMPLEX
AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY

CC WITH GRPE, THE ATPASE ACTIVITY OF DNAB (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC
 CC DNAB-LIKE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AF007813; AAC35417.1; -
 DR HSSP: U72647; AABI7396.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DNAB_1; 1.
 DR PROSITE: PS50076; DNAB_2; 1.
 DR PROSITE: PS00637; DNAB_CXXCXXG; FALSE_NEG.
 DR PRAM: PF00226; DnaJ; 1.
 KW Chapterone; DNA replication; Heat shock; Repeat.
 FT DOMAIN 3 72 DNAB-LIKE.
 FT REPEAT 150 157 CXXCXXG MOTIF.
 FT REPEAT 167 174 CXXCXXG MOTIF.
 FT REPEAT 189 196 CXXCXXG MOTIF.
 FT REPEAT 203 210 CXXCXXG MOTIF.
 SQ SEQUENCE 369 AA; 39939 MW; 408B5B11457DCDEF CRC64;

Query Match 17.7%; Score 304; DB 1; Length 369;
 Best Local Similarity 35.2%; Pred. No. 2.1e-16;
 Matches 92; Conservative 31; Mismatches 70; Indels 68; Gaps 12;

QY 4 YVEVLGVQRHASPEDIKAYRKALAKWHPDKNPENKEAEKRFQVAEAYVLSDAKRD 63
 DB 6 YYDILGYSKSANDIEKSAVKYKLAIKYHPDKNGNK-ESEKREAEVAILNDPKKQ 64
 QY 64 IYDYKRGKGLNGGGG---GSHPDSPFEFGFTFRNPDDVREFFGGDDPSFDFEDPFE 120
 DB 65 AYDDEFGAGVSGGAGGFGGAYTDFSDIFG-----DIGDFGGSSGFG----- 111
 QY 121 DFGNRRGP-RGSRSR-----GTG-SFSAFSGFP 148
 DB 111 --GGRSGPQRGSDLRKRLNLEVSLEDAALGREYKLEIRLSCVDCNCSGSKSSPITCP 168
 QY 149 SF-GSGFSSFDGTGFTSGSL-----GKGGLTSPSSSTFGSGGNGNFKSISTSTMVNGRK 202
 DB 169 DCGSGSQIRRTQGFSSVATTCPTCRGKGTIISNPCRSCGGQGLQF-KRRTINIKIPPG-- 226
 QY 203 ITRKRIYNGQERVEVEDGQ 223
 DB 226 -----VETG-SRLKVSGE 239

RESULT 10
 DNJ2.ALLPO STANDARD; PRT; 418 AA.
 AC P42824;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE DNAB PROTEIN HOMOLOG 2.
 GN LDJ2.
 OS Allium porrum (Leek).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Asparagales;
 OC Alliaceae; Allium.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bessoule J.J., Testet E., Cassagne C.;
 RT Cloning of a new isoform of a DnaJ protein from Allium porrum
 RL Epidermal cells";
 CC Plant Physiol. Biochem. 32:723-727(1994).
 -1- FUNCTION: HAVE A CONTINUOUS ROLE IN PLANT DEVELOPMENT PROBABLY

CC IN THE STRUCTURAL ORGANIZATION OF COMPARTMENTS (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: MEMBRANE-BOUND (POTENTIAL).
 CC -1- SIMILARITY: CONTAINS A DNAB-LIKE DOMAIN AND THE FOUR CXXCXXG
 CC MOTIFS FOUND IN DNAB PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X77632; CAAS4720.1; -
 DR HSSP: P25685; 1HDI.
 DR PRINTS: PR00625; DNABPROTEIN.
 DR PROSITE: PS00636; DNAB_1; 1.
 DR PROSITE: PS00637; DNAB_CXXCXXG; 1.
 DR PROSITE: PS50076; DNAB_2; 1.
 DR PRAM: PF00226; DnaJ; 1.
 DR PRAM: PF00684; DnaJ_CXXCXXG; 1.
 DR PRAM: PF01556; DnaJ_C; 1.
 KW Chapterone; Repeat; Prenylation; Lipoprotein; Multigene family.
 FT DOMAIN 11 76 DNAB-LIKE.
 FT DOMAIN 83 88 POLY-GLY.
 FT DOMAIN 81 108 GLY-RICH.
 FT REPEAT 148 155 CXXCXXG MOTIF.
 FT REPEAT 164 171 CXXCXXG MOTIF.
 FT REPEAT 191 198 CXXCXXG MOTIF.
 FT REPEAT 207 214 CXXCXXG MOTIF.
 FT LIPID 415 415 FARNESYL (BY SIMILARITY).
 SQ SEQUENCE 418 AA; 46584 MW; DCE2A4DF192329E6 CRC64;

Query Match 17.6%; Score 302.5; DB 1; Length 418;
 Best Local Similarity 32.7%; Pred. No. 3.1e-16;
 Matches 90; Conservative 25; Mismatches 73; Indels 87; Gaps 11;

QY 4 YVEVLGVQRHASPEDIKAYRKALAKWHPDK-NPENKEAEKRFQVAEAYVLSDAK 61
 DB 14 YVEVLGSKNATPEDEKKAIRKALKNHPDKGDPE-----KREIQVAEVLNDPK 66
 QY 62 RDIDYKRGKGLNGGGGSHDPSPEFGFTFRNPDDVREFFGGDDPSFDFEDPFE 121
 DB 67 REIYDYGEGGLKEKGGGGGVNDPE-----DIQSFGGG----- 104
 QY 122 FFGNRRGPGRSR-----GTG-SFSAFSGFP 145
 DB 104 -FEGGSSRGRRGRGEDVYHPLKYSLEDLYNGTSKILSLRNVLCTKCKGKSGKSGASM 162
 QY 146 GFPSF-GSGFSSFDGTGFTSGSLGKGGLTSPSSSTFGSGGNGNFKSISTSTMVNGRKIT 204
 DB 163 NCASCGSGMK-----VSIQLGFMQIQMOHQPCNECKGTGEMTSDKDRCPQCKGERV 216
 QY 205 -TRKRIYNGQERVEVE--EDGQLKSLTINGVADD 235
 DB 217 QOKKYLE-----VHYEKGQNGO--KITPPEGADE 244

RESULT 11
 MAS5_YEAST STANDARD; PRT; 409 AA.
 AC P25491;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE MITOCHONDRIAL PROTEIN IMPORT PROTEIN MAS5 (PROTEIN YDJ1).
 GN MAS5 OR YDJ1 OR YNL064C OR N2418 OR YNL2418C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomyces.
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE: 92107179.
 RA Atencio D.P., Yaffe M.P.;
 RT "MAS5, a yeast homolog of DnaJ involved in mitochondrial protein
 RT import.";
 RL Mol. Cell. Biol. 12:283-291(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 91332099.
 RA Caplan A.J., Douglas M.G.;
 RT "Characterization of YDJ1: a yeast homologue of the bacterial dnaJ
 RT protein.";
 RL J. Cell Biol. 114:609-621(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX STRAIN-S288C / FY1676;
 RX MEDLINE: 96021608.
 RA Berger P., Doignon F., Crouzet M.;
 RT "The sequence of a 44 420 bp fragment located on the left arm of
 RT chromosome XIV from *Saccharomyces cerevisiae*.";
 RL Yeast 11:967-974(1995).
 RN [4]
 RP ERRATUM.
 RX MEDLINE: 97060022.
 RA Berger P., Doignon F., Crouzet M.;
 RL Yeast 12:297-297(1996).
 RN [5]
 RP ISOPRENOLD.
 RX MEDLINE: 92406811.
 RA Caplan A.J., Tsai J., Casey P.J., Douglas M.G.;
 RT "Farnesylation of YDJ1 is required for function at elevated growth
 RT temperatures in *Saccharomyces cerevisiae*.";
 RL J. Biol. Chem. 267:18890-18895(1992).
 CC -1- FUNCTION: PROBABLY INVOLVED IN MITOCHONDRIAL PROTEIN IMPORT. IS
 CC ALSO REQUIRED FOR EFFICIENT TRANSLLOCATION OF PRE-PRO-ALPHA-FACTOR.
 CC -1- SUBCELLULAR LOCATION: CONCENTRATED IN A PERINUCLEAR RING AS WELL
 CC AS IN THE CYTOPLASM (ACCORDING TO REF. 2).
 CC -1- INDUCTION: MAS5 IS A HEAT SHOCK GENE WHOSE EXPRESSION INCREASES
 CC MODERATELY AT ELEVATED TEMPERATURES.
 CC -1- SIMILARITY: CONTAINS A DNAJ-LIKE DOMAIN AND THE FOUR CXXCGXG
 CC MOTIFS FOUND IN DNAJ PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: S74758; AAB20771.1; -;
 DR EMBL: X56560; CA93910.1; -;
 DR EMBL: U12141; AAA99647.1; -;
 DR EMBL: Z71340; CAA95937.1; -;
 DR PIR: A39659; A39659.
 DR PIR: S26703; S26703.
 DR HSSP: P25685; 1HDJ.
 DR SCD; L002503; YDJ1.
 DR PRINTS; PR00625; DNAJPROTEIN.
 DR PROSITE; PS00636; DNAJ_1; 1.
 DR PROSITE; PS00637; DNAJ_CXXCGXG; 1.
 DR PROSITE; PS50076; DNAJ_2; 1.
 DR PRAM; PF00226; DnaJ; 1.
 DR PRAM; PF00684; DnaJ_CXXCGXG; 1.
 DR PRAM; PF01556; DnaJ_C; 1.
 KW Chaperone; Heat shock; Transport; Protein transport; Mitochondrion;
 KM Repeat; Prenylation; Lipoprotein.
 FT DOMAIN 4 72 DNAJ-LIKE.
 FT REPEAT 73 103 GLY-RICH.
 FT REPEAT 143 150 CXXCGXG MOTIF.
 FT REPEAT 159 166 CXXCGXG MOTIF.
 FT REPEAT 185 192 CXXCGXG MOTIF.
 FT REPEAT 201 208 CXXCGXG MOTIF.
 FT LIPID 406 406 FARNESYL.

SQ SEQUENCE 409 AA; 44670 MW; E4539F3618DD9CF2 CRC64;
 Query Match 17.4%; Score 298.5; DB 1; Length 409;
 Best Local Similarity 29.7%; Pred. No. 6,1e-16;
 Matches 95; Conservative 36; Mismatches 86; Indels 103; Gaps 14;
 QY 4 YVEYVGVRHASPEDIKKAYRKALKMHPDKNPEKKEAEKFKFOVAAYEVLSAKRRD 63
 DB 7 FYDLIGVPTATDVEIKKAYRKALKMHPDKNP--SEEAARFKFASAYEILSPKERRD 64
 QY 64 IYDKYKREGLNMGCGGSHFDSPEFGFTFRNPDDVPFEPFGGRDPFSDFEDPEDFEF 123
 DB 65 IYDQGEDEGLGAGAGAGFPFGGFGF-----DPIFGFGA----- 102
 QY 124 GNRGRPRG-SRSRGTSFSAFSGFPSPGSGFSGFSGTSGSL-----GHGGL 172
 DB 102 GGAQRPGRPGQKDKIHEISA-----SLEELYKRTAKLALMKQILCKECPGRGK 152
 QY 173 TSF--SSTSPFGSGM-----GPFKSISTKRVNKRKITK 206
 DB 153 KGAVKKCTSCNGGQIKFYTRQMGPMIQFQTECDVCHGTGDIIDPKDRCKSCNGKKEVNE 212
 QY 207 R-IYENGQERVEVE---EDGQLKSLITGVADDDALXERRRRGQNVLPAPAGLRPPKP 262
 DB 213 RKILE-----VHEPGRMDGQ--RIVKGEAD-----QAPDVTPGD----- 247
 QY 263 PRPASLRKRPCLSKERGE 282
 DB 247 --VVEIVSERPHKSPKRRGD 264
 RESULT 12
 ID DNAJ_MYCPN STANDARD; PRT; 390 AA.
 AC P78004;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE DNAJ PROTEIN.
 GN DNAJ.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 CC Mycoplasmataceae; Mycoplasma.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 29342 / M129;
 RX MEDLINE: 97105885.
 RA Himmelfreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
 RA Herrmann R.;
 RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma*
 RT *pneumoniae*.";
 RL Nucleic Acids Res. 24:4420-4449(1996).
 CC -1- FUNCTION: INTERACTS WITH DNAJ, TO DISASSEMBLE A PROTEIN COMPLEX
 CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
 CC WITH GREP, THE ATPASE ACTIVITY OF DNAJ (BY SIMILARITY).
 CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC
 CC DNAJ-LIKE PROTEINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000015; AAB95781.1; -;
 DR HSSP; P25685; 1HDJ.
 DR PROSITE; PS00636; DNAJ_1; 1.
 DR PROSITE; PS50076; DNAJ_2; 1.
 DR PROSITE; PS00637; DNAJ_CXXCGXG; 1.
 DR PRAM; PF00226; DnaJ; 1.

```

DR PFAM: PF00684; DnaJ_CXKXGXG; 1.
KW PFAM: PF01556; DnaJ_C; 1.
RM Chapterone; DNA replication; Heat shock; Repeat.
FT DOMAIN 5 79 DnaJ-LIKE
FT REPEAT 165 172 CXKXGXG MOTIF.
FT REPEAT 183 190 CXKXGXG MOTIF.
FT REPEAT 209 216 CXKXGXG MOTIF.
FT REPEAT 223 230 CXKXGXG MOTIF.
SQ SEQUENCE 390 AA; 43167 MW; 442E1A6304E8697E CRC64;

Query Match 17.2%; Score 295; DB 1; Length 390;
Best Local Similarity 35.8%; Pred. No. 1.1e-15;
Matches 92; Conservative 24; Mismatches 67; Indels 74; Gaps 13;

QY 3 DYEEVLGVORHASPEDIKAYRKLAKMHPDKNPENKE-----EMERKKQVAEAYEVLSD 58
DB 7 DYEEVLGVSRSATADIRAKRKLAMQYHPRHKGEGFTYOKONEKREKVEAYEVLSD 66
QY 59 AKKRDIYDKYKEGLNGGGGGSHDSEPFEGFTFRNPDVFRFFPGGRDPFDF-FED 117
DB 67 TEKRMVDRFGHEGLNAGSF-----HETGF--NPFDLFNVGFG--EGSFDMOGS 113
QY 118 PFEDFEGRRRPPGRSRR-----GTG---SFF-----SAF 144
DB 114 PFDFIFS-----RGKRSQKNRVLPYDLAIVAGVDISEFEMTNGCTRTIETKKRVTSAC 168
QY 145 SGFSPFGSGGSGFPTGFTSGSL-GHGGLTSGSSTSG-----GSGMNGKSTST 195
DB 169 DGFAGEG-----ETGMVSCNSGSGNGFILKNQISFTGYVSOMLCOSGGGQKQAKHK 223
QY 196 KAVNGKRI---TTKRI 208
DB 224 KTCGSKYKKVPTTKEI 240

RESULT 13
DnaJ_BACSU
ID DnaJ_BACSU STANDARD; PRT; 372 AA.
AC P17631;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE DnaJ PROTEIN.
GN DnaJ.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / MB11;
RX MEDLINE: 92250426.
RA Weizstein M., Voelker U., Dedio J., Loebau S., Zuber U.,
RA Schlussovohl M., Hergel C., Hecker M., Schumann W.;
RT "Cloning, sequencing, and molecular analysis of the dnaK locus from
RT Bacillus subtilis.";
RL J. Bacteriol. 174:3300-3310(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S.,
RA Sato T., Takeuchi M.;
RT Submitted (May-1996) to the EMBL/GenBank/DBD databases.
RN [3]
RP SEQUENCE OF 1-24 FROM N.A.
RC STRAIN-168 / MB11;
RX MEDLINE: 90245667.
RA Weizstein M., Dedio F., Schumann W.;
RT "Complete nucleotide sequence of the Bacillus subtilis dnaK gene.";
RN [4]
RP SEQUENCE OF 257-372 FROM N.A.
RC STRAIN-168 / JH642;

```

```

RX MEDLINE: 97175542.
RA Homuth G., Masuda S., Mogt A., Kobayashi Y., Schumann W.;
RT "The dnaK operon of Bacillus subtilis is heptacistronic.";
RL J. Bacteriol. 179:1153-1164(1997).
CC -1- FUNCTION: INTERACTS WITH DnaK, TO DISASSEMBLE A PROTEIN COMPLEX
CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
CC WITH GRE, THE APASE ACTIVITY OF DnaK (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: BY HEAT SHOCK.
CC DnaJ-LIKE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M64964; AAA2529.1; -
DR EMBL: D84432; BAA12465.1; -
DR EMBL: X52064; CAA36287.1; -
DR EMBL: D83717; BAA12077.1; -
DR EMBL: 259117; CAB14488.1; -
DR PIR: S09501; S09501.
DR PIR: S27506; S27506.
DR PIR: B41874; B41874.
DR HSSP: P08622; 1XBL.
DR SUBTILIST; BG10665; DnaJ.
DR PRINTS; PR00625; DnaJPROTEIN.
DR PROSITE; PS00636; DnaJ.1; 1.
DR PROSITE; PS00076; DnaJ.2; 1.
DR PROSITE; PS00637; DnaJ_CXKXGXG; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXKXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
KW Chapterone; DNA replication; Heat shock; Repeat.
FT DOMAIN 3 71 DnaJ-LIKE.
FT DOMAIN 76 103 GLY-RICH.
FT REPEAT 145 152 CXKXGXG MOTIF.
FT REPEAT 162 169 CXKXGXG MOTIF.
FT REPEAT 188 195 CXKXGXG MOTIF.
FT REPEAT 202 209 CXKXGXG MOTIF.
SQ SEQUENCE 372 AA; 40847 MW; 353C0862F35A5A95 CRC64;

Query Match 17.1%; Score 293; DB 1; Length 372;
Best Local Similarity 33.3%; Pred. No. 1.4e-15;
Matches 90; Conservative 31; Mismatches 85; Indels 64; Gaps 10;

QY 3 DYEEVLGVORHASPEDIKAYRKLAKMHPDKNPENKEBAERKKQVAEAYEVLSDAKR 62
DB 5 DYEEVLGVSRSATADIRAKRKLAKMHPDKNPENKEBAERKKQVAEAYEVLSDAKR 62
QY 63 DYIYKVG-----KEGLNGGGGGSHDSEPFEGFTFRNPDVFRFFPGGRDPFDFEDP 118
DB 63 AHYDQFGHTDPMQGRFGGGGFGGDFG--IGF-----DILFSSIFG----- 103
QY 119 FDFFGNRRGPPGRSGRGTGFFSAFSGFPFGSGGSSFF-----DTGTFSGSLGHGCL 172
DB 103 -----GTRRRDKLRARADLYTMTLSFEDAFAFKETETIIPRETCETCGGAKPCT 157
QY 173 TSFSGSGSGMNGFKSTSTKMNCR-----KITTKRIYENG-----QERV 216
DB 158 NPEITSHCGSGQLNVEQNTPPGKVVNRRVCHNCBGTCTKITKNCADCGGKGIKKRKRI 217
QY 217 EYE-----EDGQLKSLT-----INGVAD 235
DB 218 NVTIPAGVDDGQQLRLTSGGEGINGGLPD 247

RESULT 14
DnaJ_CUCSA

```


[illegible]

ID	DNAJ_CLOAB	STANDARD;	PRT;	374 AA.
AC	P30725;			
DT	01-APR-1993 (Rel. 25, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	15-JUL-1998 (Rel. 36, Last annotation update)			
DE	DNAJ PROTEIN.			
GN	DNAJ.			
OS	Clostridium acetobutylicum.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.			
OC	Clostridium.			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE; 94123950.			
RA	Behrens S., Narberhaus F., Bahl H.;			
RT	"Cloning, nucleotide sequence and structural analysis of the			
RT	Clostridium acetobutylicum dnaJ gene.";			
RL	FEMS Microbiol. Lett. 114:53-60(1993).			
RP	[2]			
RP	SEQUENCE OF 1-72 FROM N.A.			

```

RC SEQUENCE OF 1-72 FROM N.A.
RP STRAIN-DSM 1731:
RX MEDLINE: 92250425.
RA Naderhaus F., Giebler K., Bahl H.;
RT "Molecular characterization of the dnaK gene region of Clostridium
RT aceobutylicum, including grpe_dnaJ and a new heat shock gene.";
RL J. Bacteriol. 174:3290-3299(1992).
CC -1- FUNCTION: INTERACTS WITH DNAK. TO DISASSEMBLE A PROPEIN COMPLEX
CC AT THE PHAGE LAMBDA ORIGIN OF REPLICATION. STIMULATES, JOINTLY
CC WITH GRPE, THE ATPASE ACTIVITY OF DNAK (BY SIMILARITY).
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC
CC DNAJ-LIKE PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-slb.ch/announce/
CC or send an email to license@isb-slb.ch).
CC -----
CC
DR EMBL: X69050; CAA48792.1; -.
DR EMBL: M74569; AAA23247.1; -.
DR PIR: A41873; A41873.
DR HSSP: P25685; 1HDJ.
DR PRINTS: PR00625; DNAJPROTEIN.
DR PROSITE: PS00636; DNAJ_1; 1.
DR PROSITE: PS50076; DNAJ_2; 1.
DR PROSITE: PS00637; DNAJ_CXXCXGKG; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXXCXGKG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
DR Chaperone; DNA replication; Heat shock; Repeat.
KW DOMAIN
FT 3
FT 72
FT DOMAIN
FT 76 108 GLY-RICH.
FT REPEAT 150 157 CXXCXGKG MOTIF.
FT REPEAT 167 174 CXXCXGKG MOTIF.
FT REPEAT 193 200 CXXCXGKG MOTIF.
FT REPEAT 207 214 CXXCXGKG MOTIF.
SQ SEQUENCE 374 AA; 40401 MW; 0DC6B58511F201E CRC64;

```

Query Match 16.78; Score 286.5; DB 1; Length 374;

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 11:53:56 ; Search time 53.9 Seconds

(without alignments)
424.495 Million cell updates/sec

Title: US-09-501-714-3

Perfect score: 1716
Sequence: 1 MVDYEVGVGRHASPEDIK.....EAEKGVREEVDRQSLDRT 330Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.podent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1230	71.7	241	4	075190 homo sapien
2	1158.5	67.5	242	11	054946 mus musculu
3	1060	61.8	215	4	095806 homo sapien
4	847.5	49.4	242	11	035723 mus musculu
5	363	21.2	336	5	076224 trypanosoma
6	361.5	21.1	348	10	049457 arabidopsis
7	358	20.9	338	5	076230 trypanosoma
8	343	20.0	331	5	020774 caenorhabd
9	339	19.8	348	11	089114 mus musculu
10	337	19.6	348	4	075953 homo sapien
11	332.5	19.4	330	3	013303 cryptococu
12	332	19.3	178	4	013431 homo sapien
13	330	19.2	392	2	084345 chlamydia t
14	327	19.1	354	5	096123 plasmodium
15	309	18.0	402	5	045502 caenorhabd
16	307.5	17.9	331	10	09X115 arabidopsis
17	303.5	17.7	392	2	092999 chlamydia p
18	298	17.4	411	5	096455 dictyostel
19	295.5	17.2	420	10	092MK3 salix gligl
20	291	17.0	369	2	025890 helicobacte

ALIGNMENTS

RESULT	ID	075190	PRELIMINARY;	PRT;	241 AA.
AC	075190	075190			
DC	01-NOV-1998	(TREMblrel. 08, Created)			
DT	01-NOV-1998	(TREMblrel. 08, Last sequence update)			
DT	01-NOV-1999	(TREMblrel. 12, Last annotation update)			
DE	MSJ-1.				
GN	MSJ-1 OR HSJ2.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;				
OC	Eukarya; Primates; Catarrhini; Hominoidea; Homo.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=TESTIS;				
RA	SAITO T., SEKI N.;				
RT	"A new member of human dnaJ-related gene family."				
RL	Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	ZHANG W., WAN T., YUAN Z., CAO X.;				
RT	"HSJ2, a novel human homologue of the bacterial heat-shock protein DnaJ".				
RL	Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AB014888; BAA32209.1; -.				
DR	EMBL; AF075601; AAD43194.1; -.				
DR	HSSP; P25685; 1HDJ.				
DR	PROSITE; PS00636; DnaJ_1; 1.				
DR	PFAM; PF00226; DnaJ; 1.				
KW	Heat shock.				
SO	SEQUENCE	241 AA;	26900 MW;	D2F7F6BF CRC32;	

Query Match 71.7%; Score 1230; DB 4; Length 241;
Best Local Similarity 100.0%; Pred. No. 1.1e-93;
Matches 231; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MVDYEVGVGRHASPEDIKAKYRLAKWHPDKNPKKEAEKRFKOVAAAYEVLSPAK 60	096212 plasmodium
Db	1	MVDYEVGVGRHASPEDIKAKYRLAKWHPDKNPKKEAEKRFKOVAAAYEVLSPAK 60	087778 lactobacill
Qy	61	KRDYDKKRGKGLNGGGGGSHFSPPEFGFTFRNPDDVFRPFGGRDPFDFDPDFE 120	092JQ2 helicobacte
			092S26 hevea brasl
			042530 arabidopsis
			022663 arabidopsis
			043177 solanum tub
			09Xc66 porphyromon
			092re0 nicotiana t
			09Y899 schizosacch
			094625 schizosacch
			092f65 methylovoru
			065160 zea mays (m
			092wz3 thermotoga
			068797 campylobact
			024074 medicago sa
			052065 pasteurella
			087385 vibrio harv
			075472 homo sapien
			026953 trypanosoma
			09Xc52 rhodothermu
			074752 schizosacch
			018427 geodia cydo
			077048 plasmodium
			085213 campylobact

```

Db 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 120
QY 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 180
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 180
QY 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231
    |||
Db 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231

RESULT 2
054946 PRELIMINARY; PRT; 242 AA.
ID 054946
AC 054946;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MRJ.
GN MRJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-129; TISSUE-TROPHOBLAST;
RA HUNTER P.J., SPANSON B.J., HAENDEL M., LYONS G.E., CROSS J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF035962; AAC16759.1; -.
DR HSSP: P25685; 1HDJ.
DR PROSITE: PS00636; DNAJ_1; 1.
DR PFM: PF00226; DnaJ; 1.
SQ SEQUENCE 242 AA; 27012 MW; AC627078 CRC32;

Query Match 67.5%; Score 1158.5; DB 11; Length 242;
Best Local Similarity 93.5%; Pred. No. 8.2e-88;
Matches 217; Conservative 5; Mismatches 9; Indels 1; Gaps 1;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
QY 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 119
    |||
Db 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 120
QY 120 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 179
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 180
QY 180 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 231
    |||
Db 181 GSGGMGNFKSISTSTKMVNGRKITTKRIYENGOERVEVEDQGLKSLTING 232

RESULT 3
095806 PRELIMINARY; PRT; 215 AA.
ID 095806
AC 095806;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNAJ-LIKE 2 PROTEIN.
GN HSJ2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-TESTIS;
RX MEDLINE: 99115663.
RA PEI L.;

```

```

RT "Plutitary tumor-transforming gene protein associates with ribosomal
RT protein S10 and a novel human homologue of DnaJ in testicular cells.";
RL J. Biol. Chem. 274:3151-3158(1999).
DR EMBL: AF080569; AAD16010.1; -.
DR HSSP: P25685; 1HDJ.
DR PROSITE: PS00636; DNAJ_1; 1.
SQ SEQUENCE 215 AA; 24051 MW; 49385BFB CRC32;

Query Match 61.8%; Score 1060; DB 4; Length 215;
Best Local Similarity 97.5%; Pred. No. 8.7e-80;
Matches 197; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
QY 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 120
    |||
Db 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 120
QY 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 180
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 180
QY 181 GSGGMGNFKSISTSTKMVNGRK 202
    |||
Db 181 GSGGMGNFKSISTSTKMVNAEK 202

RESULT 4
035723 PRELIMINARY; PRT; 242 AA.
ID 035723
AC 035723;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE TESTIS SPECIFIC DNAJ-HOMOLOG.
GN HSJ3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE: 98189155.
RA BERRUTI G., PEREGO L., BORGONOVO B., MARTEGANI E.;
RT "MSJ-1, a new member of the DnaJ family of proteins, is a male germ
RT cell-specific gene product.";
RL Exp. Cell Res. 239:430-441(1998).
DR EMBL: U95607; AAC13944.1; -.
DR HSSP: P25685; 1HDJ.
DR MGD: MGI:1306822; HsJ3.
DR PROSITE: PS00636; DNAJ_1; 1.
DR PFM: PF00226; DnaJ; 1.
SQ SEQUENCE 242 AA; 26678 MW; D3DCA2B7 CRC32;

Query Match 49.4%; Score 847.5; DB 11; Length 242;
Best Local Similarity 70.0%; Pred. No. 2.8e-62;
Matches 163; Conservative 30; Mismatches 37; Indels 3; Gaps 3;

QY 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
    |||
Db 1 MVDYEVVLGVORHASPEDIKAYRKALAKWHPDKNPENKEAEERKQVAEAYEVLSDAK 60
QY 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 118
    |||
Db 61 KRDIYKYGKGLNGGGGSHFDSPEFGFTFRNPDDVFRFFGGDRDPFSDFEDDPPE 120
QY 119 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 178
    |||
Db 121 DFFGNRRGPRGSRSGTGSFFSAFSGFSPSGSFTGTSGSLGHGGLTSSSTSF 179

```


DR PFAM; PF01556; DnaJ_C; 1.
KM Heat shock.
SQ SEQUENCE 348 AA; 39133 MW; 92AF8A03 CRC32;

Query Match 19.6%; Score 337; DB 4; Length 348;
Best Local Similarity 32.2%; Pred. No. 3.8e-20;
Matches 96; Conservative 39; Mismatches 101; Indels 62; Gaps 11;

QY 3 DYIVVLGVORHASPEDIKKAYRKIALKWHDPKNPENKEEAERKFKOVAEAYEVLSDAKK 62
D 4 DYVKILIPSGANDEIKKAYRKMAKLYHPDKNE--PNAEEKKEIAEAYDVLSDPKR 61
QY 63 DYVKKRKEGNGGG--GSHFDSPEFEGTFR-NPDVYREFEGGDPSPFEDEPF 119
D 62 GLVDQYEGEGTKTGSG--SSGSFHTFHPDPATPAASFFGG-----SNPF 108
QY 120 EDFGNGRPGSGSRGSGFSSFGSGFSSFGTFTGSLGHGILTSFSTS 179
D 109 DIFPASSRTR-----PFGSDPDMDVDEDEDFGAFGRGFGNLSRGPRRA 156
QY 180 FGSGMGNFSTSTSTKMNKRIKTIYENGQERVEVE--DGLKSLTI----- 230
D 157 -----PEPLPRKRYODPPVYH--ELRVSLREIYHSGTRKMKITRRRLNPD 200
QY 230 -NGVADDDALXEEBMRG-----QNVLPAPAGLPRPPRPAALIRKXPHCLSKKEG 281
D 201 GRVTRTEDKILHIVIKRGWKEGRTITEPEKGDATPDNIPADIVELDKPHAFRRRDG 258

RESULT 11

ID 013303 PRELIMINARY; PRT; 330 AA.
AC 013303;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE STS1 PROTEIN.
GN STS1.
OS Cryptococcus curvatus.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Tremellales;
OC anamorphic Tremellales; Cryptococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ATCC 20509;
RX MEDLINE; 98220306.
RA SPECHT V., LUBECK M., KINDL H.;
RT "Heat shock transiently enhances the synthesis rate of Sis1p, a
RT ribosome-associated DnaJ protein in the oleagenous yeast *Apilotrichum*
RT curvatum";
RL Yeast 14:419-430(1998).
DR EMBL; Y12080; CAA72798.1; -.
DR HSSP; P25685; 1HDJ.
DR PROSITE; PS00636; DNaJ_1; 1.
DR PFAM; PF00226; DnaJ; 1.
SQ SEQUENCE 330 AA; 34596 MW; 4FD216DA CRC32;

Query Match 19.4%; Score 332.5; DB 3; Length 330;
Best Local Similarity 32.8%; Pred. No. 8.3e-20;
Matches 97; Conservative 33; Mismatches 78; Indels 89; Gaps 13;

QY 3 DYIVVLGVORHASPEDIKKAYRKIALKWHDPKNPENKEEAERKFKOVAEAYEVLSDAKK 61
D 6 EYVYTLISKRSADIKKAYRKESLKWHPDKNGDKRATAEKKFKGAEAYEVLSDPEK 65
QY 62 RDIDYKXGKBSLN-----GGGGGSHDSDPEFG-----ETFRPDVDFEFEGGR 107
D 66 RKIIDQGEERLKGMPAGGGGGPGFSS--FGAGAAPASPTPTDNDIFNAFF-- 120
QY 108 DPFSDFEDEPFDFGNGRGRGSRSGTGSFSAF--SGPFSFGSGFSSFDGFTGFTSG 165
D 120 -----SQMGSGGGGGGDEYAFSGCGGPGGGRARRRGRGCMG 159

QY 166 SL-GHGILTSSTSFSGSG-----MGNFKSSTS 194
D 160 GMGGMGMPGMG--GMGGMGMPGPGYARPEPTPAVETIERPLPLSLELYGGTRLRIT 217
QY 195 TKMVGKRIKTIYENGQERVEVEEDGL-KSLTINGVADDDALXEEBMRGONV 249
D 218 RHMNRGR--TNEKILE-----VAIKPGWAKGTKYTFAGAGNED-----EYRGONI 261

RESULT 12

ID 013431 PRELIMINARY; PRT; 178 AA.
AC 013431;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HEAT SHOCK PROTEIN HSP40 HOMOLOG.
GN DNaJw.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER.
RA WON M., MOON K.M., LEE C.E., YOO H.S.;
RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.
DR EMBL; U40992; AAC14483.1; -.
DR EMBL; U41290; AAB07346.1; -.
DR HSSP; P25685; 1HDJ.
DR PROSITE; PS00636; DNaJ_1; 1.
DR PFAM; PF00226; DnaJ; 1.
KW Heat shock.
SQ SEQUENCE 178 AA; 19930 MW; 5A1337E4 CRC32;

Query Match 19.3%; Score 332; DB 4; Length 178;
Best Local Similarity 49.4%; Pred. No. 4.2e-20;
Matches 76; Conservative 16; Mismatches 40; Indels 22; Gaps 6;

QY 3 DYEVTLGVORHASPEDIKKAYRKIALKWHDPKNPENKEEAERKFKOVAEAYEVLSDAKK 62
D 4 DYICLTIGEKASDEDIKKAYRKQALKRHPDKN--KSQAEEKKEVAAEAYEVLSDPKR 61
QY 63 DYDKYKREGNGGGGSHDSDPEFEGTFRNPDVYREFEGGDRPSPFEDEPFEDF 122
D 62 EYIDQGEERLKGAGGAGTDDGGGTFRTYF-HGDPHATPAAFKXGSPFXI-----F 111
QY 123 FGNRRGPRGSRG-----TGSFSAF-----SGFP 148
D 112 FGRRWG--GGRDSEMEIDGDFSAFGFSMGYP 143

RESULT 13

ID 084345 PRELIMINARY; PRT; 392 AA.
AC 084345;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE HEAT SHOCK PROTEIN J.
GN DNaJ.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/WM-3/CX;
RA STEPHENS R.S., KALMAN S., LAMMEL C.J., PAN J., MARATHE R., ARAYIND L.,
RA MITCHELL W.P., OLINGER L., TATUSOV R.L., ZHAO Q., KOONIN E.V.,
RA DAVIS R.W.;
RT "Genome Sequence of an Obligate Intracellular Pathogen of Humans:
RT Chlamydia trachomatis";
RL Science 0:0-0(1998).

RN SEQUENCE FROM N.A.
 RP [2]
 RC STRAIN-D/UM-3/CX;
 RA STEPHENS R.S., KALMAN S., LAMMEL C.J., FAN J., MARATHE R., ARAVIND L.,
 RA MITCHELL W.B., LINGER L., TATISOV R.L., ZHAO Q., KOONIN E.V.,
 RA DAVIS R.W.;
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE001507; AAC67936.1; -.
 DR HSP: P08622; 1XBL.
 DR PROSITE; PS00636; DnaJ_1; 1.
 DR PFAM; PF00226; DnaJ; 1.
 DR PFAM; PF01556; DnaJ_C; 1.
 DR PFAM; PF00684; DnaJ_CXXCGXG; 1.
 DR PRINTS; PR00625; DNAJPROTEIN.
 KW Heat shock.
 SQ SEQUENCE 392 AA; 41916 MW; 8D33F656 CRC32;

```

QY 2 VDEVEYLGRHNSPEIDIKKAYRKILKMKHPDKNPNENKKEAEKRFQVQAEVEYLSADKK 61
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||: 11:|
Db 1 MDYITLIGVAKTAPEIKKAIKRIKLAVKYIHPDKNP-GDAEAEKRFQVSEALVEYLGDAOK 59
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 62 RDIDYKYGKBEGLNGCGGGGSHFDSPEFCGTFEPNPDVFERFFGGADPFSF-----F 114
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 60 RESYDRGKCKGPFPGAGCGFG-----GAGMGMMDALRTFMGA---FGGPGFGCGGCF 108
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 115 FEDPFE-----EDFEGRRRPRRSRSG-----TGSFSAFSGPPS--FGSGFSPFGMFTS 163
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 109 FEGFLGGLGEFPGKRGSGESSRQGAOKKHYITLTFEAAKGYEKELLYSGYRSCD----- 164
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
QY 164 FGSJGHGGLTFSFSSTFG-----GSGN-----GNFSISTSTMV-NGKRITTKRIY 209
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 164 -ACGSGS-----ANTAGVYVCDRCCKSGCGVYVSREFFSMASTCPGCGSEGRVITDPCSV 217
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 210 ENGGERVEVEDGQLSLTITNGVADDDALKEEMR-----RGQVVLPA 252
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 218 CRGGRIK-----DKRSVYHNIPGVDVSGMRLEKMEGTGADAGQAGAPA 259
   :|||:||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

RESULT	14	
096123		
ID	096123	PRELIMINARY;
AC	096123;	PRT; 354 AA.
DT	01-MAY-1999 (TREMBLrel. 10, Created)	
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)	
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)	
DE	PROTEIN WITH DNAJ DOMAIN.	
GN	PF00909C.	
OS	Plasmodium falciparum	
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 99021743.	
RA	GARDNER M.J., TETTELIN H., CARUCCI D.J., CUMMINGS L.M., ARAYIND L.,	
RA	KOONIN E.V., SHALLOO S., MASON T., YU K., FUJII C., PEDERSON J.,	
RA	SHEN K., JING J., ASTON C., LAI Z., SCHWARTZ D.C., BETTEA M.,	
RA	SAIIZBERG S., ZHOU L., SUTTON G.G., CLAYTON R., WHITE O., SMITH H.O.,	
RA	FRASER C.M., ADAMS M.D., VENTER J.C., HOFFMAN S.L.;	
RT	"Chromosome 2 sequence of the human malaria parasite Plasmodium	
RT	falciparum.";	
RL	Science 282; 1126-1132(1998).	
DR	EMBL; AE001370; AAC71808.1; -.	
DR	HSP; P08623; 1XB1.	
DR	PROSITE; PS00636; DNAJ_1; 1.	
Q0	SEQUENCE 354 AA; 40329 MW; 35500567 CRC32;	

Query Match	19.18;	Score 327;	DB 5;	Length 354;
Best Local Similarity	37.68;	Pred. No. 2.6e-19;		

[illegible]

RESULT	15	
ID	045502	
AC	045502	PRELIMINARY;
DT	01-JUN-1998	(TREMBLrel, 06, Created)
DT	01-JUN-1998	(TREMBLrel, 06, last sequence update)
DT	01-NOV-1999	(TREMBLrel, 12, last annotation update)
DE	F39B2.10	PROTEIN.
GN	F39B2.10.	
OS	Caenorhabditis elegans.	
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabdillia; Rhabdillida;	
OC	Rhabdillina; Rhabdilloidea; Rhabdillidae; Peloderinae; Caenorhabditis.	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	DOBSON R.;	
RL	Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE; 94150718.	
RA	WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERS M.,	
RA	BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,	
RA	CRAWTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,	
RA	GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,	
RA	JONES M., KESHAW J., KIRSTEN T., LAISTER N., LABRIELLE P.,	
RA	LIGHTNING J., LLOYD C., MCMBRAY A., MORIMORE B., O'CALLAGHAN M.,	
RA	PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,	
RA	SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULTON J.,	
RA	THIERRY-MIEG J., THOMAS K., VADDIN M., VAUGHAN K., WATERSTON R.,	
RA	WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;	
RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.	
RT	elegans.";	
RL	Nature 368:32-38(1994).	
CC	-1. SMILARIITY: TO OTHER PROKARYOTIC DNA, AND TO EUKARYOTIC DNA-LIKE	
CC	PROTEINS.	
DR	EMBL; Z29834; CAB07390.1; -.	
DR	HSSP; P25685; 1HDU.	
DR	PROSITE; PS00636; DNa1_1; 1.	
DR	PROSITE; PS00637; DNa1_CXXCXGXG; 1.	
DR	PFAM; PF00226; Dna1; 1.	
DR	PFAM; PF00684; Dna1_CXXCXGXG; 1.	
DR	PFAM; PF01556; Dna1_G; 1.	
DR	PRINTS; PR00625; DNa1PROTEIN.	
KW	Chaperone; DNA replication.	
SO	SEQUENCE 402 AA; 44308 MW; 592PB9CF CRC32;	

Query Match Similarity	18.0%	Score 309;	DB 5;	Length 402;
Best Local Similarity	29.4%	Pred. No. 9,1e-18;		
Matches	91;	Conservative	38;	Mismatches 64;
				Indels 116;
				Gaps 11.
QY	4	YVEVLGVQVHASPEDIKKAYRKRLAKMHPDDKPNKEEKERFKOVAAEYVLSDAKKDD	63	
DB	7	YVDVGVKFRDADSNELKKAYRKRMALKEHDDKPNDDGEQ---FKDISQAEYVLSDEKKQ	62	


```

OY      64 IYDKRGKGLNGGGGSHDPSPEGFETRNPDVREFPFGGDPRSPFFEDDPDEDF 123
        ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      63 IYDDGGEGALDGGGGGGGBGNFP-----DVDFMFG-----95
OY      124 GNRRCPGRSRSGTGSFSPASGTFSPSGSFTGFTSGSLGHGLT----SFST 178
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      95 GGGRGGRGEREVK-----PTVHNIRVTLDITLYK-----GYTKKLISRTAT 135
OY      179 SFGSGGMNPFSTSTRKMNVRKITTRIVE-----211
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      136 CKCKGKGLAGNEGSKKECSDCHGRKITX-RYIRMGPMYQOMQSHCDSCNGEBSFLEKDRC 194
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      211 ---NGSERVEEE-----DGOLKSILTTINGVAD-----DDALXEB 241
        ||| :: | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      195 KCKGKKRVKVEDLIEVGITPGMKDGE--KFVEGKGDEYIGLEPKDGVVLDEVENHEK 252
        | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY      242 RMRGRONVL 250
        | : | | | : |
Db      253 FVRKGDNI L 261

```

Search completed: June 30, 2000, 11:53:57
Job time: 7114 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 30, 2000, 09:55:23 ; Search time 53.9 Seconds
(without alignments)
460.312 Million cell updates/sec

Title: US-09-501-714-1
Perfect score: 1883
Sequence: 1 MAPGNLSFPCLLLYLIGAV.....IKQLLKQSGVQKVYINGLOGY 358

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 225878 seqs, 69334122 residues

Total number of hits satisfying chosen parameters: 225878

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SPTREMBL_12:*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	559	29.7	412	4	060884	060884 homo sapien
2	553	29.4	412	11	035824	035824 rattus norv
3	552	29.3	411	5	096455	096455 dictyostell
4	551	29.3	420	10	042530	042530 arabidopsis
5	551	29.3	420	10	022663	022663 arabidopsis
6	534	28.4	419	10	043177	043177 solanum tub
7	526	27.9	420	10	0926K3	0926K3 salix gilgt
8	515	27.3	419	10	065160	065160 zea mays (m
9	509.5	27.1	415	4	014711	014711 homo sapien
10	508	27.0	415	10	0926Z6	0926Z6 hevea brasl
11	507.5	27.0	423	10	024074	024074 medicago sa
12	495	26.3	439	5	016303	016303 caenorhabdi
13	490	26.0	376	2	066921	066921 aquifex aeo
14	470.5	25.0	331	5	020774	020774 caenorhabdi
15	469	24.9	371	2	0926F5	0926F5 methylovoru
16	457	24.3	399	5	026952	026952 trypanosoma
17	457	24.3	407	3	074752	074752 schizosacch
18	455.5	24.2	390	5	026953	026953 trypanosoma
19	450.5	23.9	348	11	089114	089114 mus musculu
20	448.5	23.8	348	4	075953	075953 homo sapien

21	441	23.4	328	5	096212	096212 plasmodium
22	439	23.3	354	5	096123	096123 plasmodium
23	438.5	23.3	379	2	008356	008356 rhodopseude
24	436.5	23.2	377	2	032465	032465 buchnera sp
25	436.5	23.2	385	2	087385	087385 vibrio harv
26	431.5	22.9	348	10	049457	049457 arabidopsis
27	429	22.8	402	5	045502	045502 caenorhabdi
28	426.5	22.7	408	5	043978	043978 babesia bov
29	417.5	22.2	377	2	055505	055505 synecocyst
30	410	21.8	331	10	09XIF5	09XIF5 arabidopsis
31	406	21.6	392	2	0929E9	0929E9 chlamydia p
32	405.5	21.5	368	2	069269	069269 bacillus sp
33	403	21.4	369	2	09WZV3	09WZV3 thermotoga
34	402.5	21.4	389	1	034135	034135 halobacteri
35	401.5	21.3	412	5	018427	018427 geodia cydo
36	401	21.3	392	2	084345	084345 chlamydia t
37	398.5	21.2	379	2	052065	052065 pasteurella
38	392.5	20.8	498	10	P92424	P92424 pisum sativ
39	390	20.7	383	2	09XCA6	09XCA6 porphyromon
40	389.5	20.7	378	2	049762	049762 mycobacteri
41	387.5	20.6	382	2	005825	005825 mycobacteri
42	385	20.4	316	2	09XCB2	09XCB2 rhodothermu
43	382	20.3	383	2	087778	087778 lactobacilli
44	378.5	20.1	379	2	052164	052164 streptomyce
45	377.5	20.0	413	3	094657	094657 schizosacch

ALIGNMENTS

```
RESULT 1
ID 060884 PRELIMINARY; PRT; 412 AA.
AC 060884;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNAJ PROTEIN.
GN HIRIP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA LORAIN S., BRENDEN C., SCAMPS C., LECUSE Y., LIPINSKI M.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE
CC PROTEINS.
DR EMBL: AJ001309; CA04669.1; -.
DR EMBL: Y13350; CA73791.1; -.
DR HSSP: P25685; 1HDJ
DR PROSITE: PS00636; DNAJ_1; 1.
DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
DR PRINTS: PRO0625; DnaJPROTEIN.
DR Chapterone: DNA replication; Repeat.
KW SEQUENCE 412 AA; 45745 MW; F8BF3B0 CRC32;
SQ
```

Query Match 29.7%; Score 559; DB 4; Length 412;

Best Local Similarity 37.9%; Pred. No. 5.7e-38;

Matches 130; Conservative 56; Mismatches 119; Indels 38; Gaps 10;

```
QY 19 AVIGRDFYKILGVPRASIKIRKAYRKLAQLQHPDPDPDPOAQERFODLGAAYEVLS 78
I: | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 2 ANVADTKLYDILGVPASASENELKKAYRKLAKEHPDKNPN--AGDKFKEISRAYEVLS 58
```


SO SEQUENCE 419 AA; 46700 MW; 66DB838A CRC32;

Query Match 28.4%; Score 534; DB 10; Length 419;
Best Local Similarity 36.7%; Pred. No. 6.5e-36;
Matches 130; Conservative 66; Mismatches 112; Indels 44; Gaps 13;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPQAOEKFQDLAGAAYEVLSDSEKQ 85
D 14 YEILGVPKTAQEDLKKAYRKALIKNHPDKGDD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----DIFSFFG--DFGFMFGTTPROQDRNIPGSDIIVDL 138
D 70 YDQGEALKEGSGGGGSHDPDIFSSFFGSPFG---GGGSSSRGRGRGDEYVHPL 126
QY 139 EYTLVEYVAGNF--VEVVRN-----KVARQAPGRKRCN--CR-QEMRTT--OLGPGRFQ 186
D 127 KVSLEDLYNGSKSLSLSRNVLSCKGKSGKSGASKSGSCGCGSGMKVYTRQDGPMTIQ 186
QY 187 MTQ-----EVV-----CDECPNVKLVNEERTLEVELEPGVRDGMEXEYPIGEGEPH 231
D 187 QMHPKNECKGTGEMINDKDCGCGCKEYVQEKVLEVVEYVEKGMQNGKITFPFGADEA 246
QY 232 VDGEPPGLRFRIRIKYVKKHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVIH--SR 289
D 247 PDVTVGDIVFLQCKHDPKFRKRGDDLFVEHTLSLTALCGFQVILHLDROGLLISQPG 306
QY 290 DKITRPGAXXMKKEGLPNFDNNNIKSLITFDPVDPKQDLEAREGIRKOLL 343
D 307 GEVVKPDQFKAINDEGMVQRPFRMRGKLYIHFTVERP--DLSLSPDMCKALEAVLPRAVQ 359

RESULT 7
Q92WK3 PRELIMINARY; PRT; 420 AA.
AC Q92WK3
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE DNAB_HOMOLOG.
GN SGJ3.
OS Salix gligiana.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eurosids I; Malpighiales; Salicaceae; Salix.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=FLOWER;
RA FUTAMURA N., ISHII-MINAMI N., HAYASHIDA N., SHIMOHARA K.,
RT molecular chaperone DnaJ and DnaK in the Japanese willow (Salix
gligiana Seemen).
RT Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL -1- SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC DNAB-LIKE
CC PROTEINS.
CC EMBL: AB015601; BAA35121.1; -.
DR HSSP: P25685; 1HDJ.
DR MENDEL: 35038; Salgi.1253;35038.
DR PROSITE: PS00637; DNAB_1; 1.
DR PROSITE: PS00637; DNAB_CXXCXGXG; 1.
KW Chaperone; DNA replication.
SQ SEQUENCE 420 AA; 46563 MW; 84A45244 CRC32;

Query Match 27.9%; Score 526; DB 10; Length 420;
Best Local Similarity 35.8%; Pred. No. 2.9e-35;
Matches 129; Conservative 65; Mismatches 124; Indels 42; Gaps 12;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPQAOEKFQDLAGAAYEVLSDSEKQ 85
D 14 YVEVLGVSASQDDLKAYRKALIKNHPDKGDD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----SHG--DIFSFFGDFGFMFGTTPROQDRNIPRSDIIVDL 139

D 70 YDQGEALKEGSGSGGSHDPDIFQSFPG--GNPFGGSGSSRGRGRKRGEDVTHPLK 128
QY 140 VTLVEYVAGNF--VEVVRN-----KVARQAPGRKRC--NCR--QEMRTTOLGPGRFQ 187
D 129 VSEFDLYNGTSKSLSLSRNVLSCKGKSGKSGASKSGSCGCGSGMKVSIHRLGSPMTIQ 188
QY 188 TOEVV-----CDECPNVKLVNEERTLEVELEPGVRDGMEXEYPIGEGEPH 232
D 189 QMHPKNECKGTGETINDKRCPOCKGKGVQEKVLEVVEYVEKGMQNGKITFPFGADEAP 248
QY 233 DGEPPGLRFRIRIKYVKKHPIFERRGDDLYTNVTVSLVESLVGFEMDITHLGHKVIHSD-- 291
D 249 PDVTVGDIVFLQCKHDPKFRKRGDDLFVEHTLSLTALCGFQVILHLDROGLLISQPG 308
QY 291 KITRPGAXXMKKEGLPNFDNNNIKSLITFDPVDPKQDLEAREGIRKOLL-KQGSVO 349
D 309 EVVVKPDQFKAINDEGMVQRPFRMRGKLYIHFTVERP--DLSLSPDMCKALEAVLPRAVQ 367

RESULT 8
O65160 PRELIMINARY; PRT; 419 AA.
ID O65160
AC O65160;
DT 01-AUG-1998 (T-EMBLrel. 07, Created)
DT 01-AUG-1998 (T-EMBLrel. 07, Last sequence update)
DT 01-NOV-1999 (T-EMBLrel. 12, Last annotation update)
DE DNAB-RELATED PROTEIN ZMDJ1.
GN ZMDJ1.
OS Zea mays (Maize).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
OC Poaceae; Zea.
RN [1]
RP SEQUENCE FROM N.A.
RA BASZCZYNSKI C.L., BARBOUR E., ZEKA B., MADDOCK S.E., SWENSON J.L.;
RL Maydis 42:189-201(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA BASZCZYNSKI C.L.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER PROKARYOTIC DNAB, AND TO EUKARYOTIC DNAB-LIKE
CC PROTEINS.
CC EMBL: AF053468; AAC08009.1; -.
DR HSSP: P25685; 1HDJ.
DR MENDEL: 29613; Zeama.1253;29613.
DR PROSITE: PS00637; DNAB_CXXCXGXG; 1.
DR PROSITE: PS00637; DNAB_1; 1.
DR PFAM: PF00226; DnaJ; 1.
DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
DR PFAM: PF01556; DnaJ_C; 1.
DR PRINTS: PR00625; DNABPROTEIN.
KW Chaperone; DNA replication.
SQ SEQUENCE 419 AA; 46725 MW; BE97EBB2 CRC32;

Query Match 27.3%; Score 515; DB 10; Length 419;
Best Local Similarity 35.8%; Pred. No. 2.3e-34;
Matches 125; Conservative 61; Mismatches 115; Indels 48; Gaps 10;

QY 26 FYKILGVPKRSIKDKKAYRKALQLHPDRNPDDPQAOEKFQDLAGAAYEVLSDSEKQ 85
D 14 YEILGVPKTAQEDLKKAYRKALIKNHPDKGDD---PEKFKELAQAYEVLSDPEKREI 69
QY 86 YDTGEEGLKRGHSSHC-----GDISSHFG--DFGFMFGTTPROQDRNIPGSDIIVDL 139
D 70 YDQGEALKEGSGGGSHDPDIFSSFFGSPFG---GGGSSSRGRGRGDEYVHPLK 126
QY 140 VTLVEYVAGNFVEYVVRNPNV-----ARQAPGRKRCNCRQEMRTTOLGPGRF 185
D 127 VSLLEDLYNGTSKSLSLSRNVLSCKGKSGKSGASKSGSMRCGCGSGMKVYTR--QLGSPMT 184
QY 186 QMTQEVVDCDECPNV-----KLVNEERTLEVELEPGVRDGMEXEYPIGEGE 229

```
Db 185 QOMQO-PCNECKGTGESINEKDRCPGCKGEKVEIDEKTVLEHVHEKGMQHNKITEFGAD 243
QY 230 PHVDGEGDLFRIRKVVKHPFERRGDDLYTNVTVSLVESLVGEMDITHDGHKVIISR 289
Db 244 EAPDTVGTGDIYFVQKSHSKFKRKGEDLFYEHTLSLEALCGFQFVTLHDNRLLIKS 303
QY 290 D-ITRPGAXXMKKGESLPFNDNNIKGSLITFDVDFPEKQLEEAR 336
Db 304 DGEVAVKDFDKAINDGMPYQRPFKMGKLYIHFTVEFPDSLAPDOCK 352

RESULT 9
ID 014711 PRELIMINARY; PRT; 415 AA.
AC 014711;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNU3/CPB3.
GN DNU3/CPB3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA EDWARDS M.C., LIEGROIS N., HORECKA J., DEPINHO R.A., SPRAGUE G.F.,
RA TYERS M., ELLEDGE S.J.;
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNUJ, AND TO EUKARYOTIC DNUJ-LIKE
CC PROTEINS.
DR EMBL; AF011793; AAB69313.1; -.
DR HSSP; P25685; 1HDJ.
DR PROSITE; PS00637; DNUJ_1; 1.
DR PROSITE; PS00637; DNUJ_CXKXG; 1.
DR PRAM; PF00226; DnuJ_1.
DR PRAM; PF00684; DnuJ_CXKXG; 1.
DR PRAM; PF01556; DnuJ_C; 1.
DR PRINTS; PR00625; DNUJPROTEIN.
KW Chapterone; DNA replication.
SQ SEQUENCE 415 AA; 46306 MW; 11B8B5D7 CRC32;

Query Match 27.1%; Score 509.5; DB 4; Length 415;
Best Local Similarity 37.0%; Pred. No. 6.4e-34;
Matches 128; Conservative 55; Mismatches 122; Indels 41; Gaps 12;

QY 19 AVIGRDPYKILIGVPRASIKDKKAYRKALQLHPDRNPDPQAOE-KFODLAGAYEVL 77
Db 2 ANVADTKLIDLVGPAGASENELKAYRKLAKEVHPKN---PQMETNFEKISFAYEVL 58
QY 78 SDSEKROYDTYGEGLKDGHOSS--HGDIFF--HFGDFGFMGCTPRQODRNIPRGS 132
Db 59 SNPKRELYDRYGGGLGEGSGGGGGLIFSLVFCGGLGFM--GNQSSRNRRRGE 116
QY 133 DIYDLEVTLEEVYAGNVEVYVARKVARQA---PGK---RCN-CR---QEMRTQL 180
Db 117 DMHPPLKVSLELDLYNGKTTKLQLSKNVLCASCSGGSGGAGVQKCSACRGKGVRLIMIQ 176
QY 181 GPRGFOMTQEVV-----CDECPNVKLVNEERTLEVEIPGVDRGMEYEPFI 225
Db 177 APGAVQOMQOSVSCDNGCEVINEKDRCKKCEGKVIKEVILVHVKGKMGQRITFT 236
QY 226 GEGEPHVDGEPGLFRIRKVVKH-PIFERRGDDLYTNVTVSLVESLVGFEMDITHDGHK 284
Db 237 GEAQDAPRMEWETLFFLLPGKMEVVFQROGNDLHMTYKIGLVEALCGFQFVTLSDGRQ 296
QY 285 VHS--RDKITRPGAXXMKKGESLPFNDNNIKGSLITFDVDFPK 328
Db 297 IVVKKPGKVIIEPGCVRVVRGEGMPQYRNPFKGGTLKFDVQFPE 342

RESULT 10
```

```
09ZS26
ID 09ZS26 PRELIMINARY; PRT; 415 AA.
AC 09ZS26;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNUJ PROTEIN.
OS Hevea brasiliensis (Para rubber tree).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids I; Malpighiales; Euphorbiaceae;
OC Hevea.
RN [1]
RP SEQUENCE FROM N.A.
RA SHIN D.H., YANG J., CHOW K.-S., HAN K.-H.;
RA "A Hevea brasiliensis homolog of DnuJ (HvDnuJ) is Expressed Abundantly
RA in latex and induced by Ethylene (Accession No. AF085275) (PGR99-
RA 004)."
RL Plant Physiol. 119:363-363(1999).
CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNUJ, AND TO EUKARYOTIC DNUJ-LIKE
CC PROTEINS.
DR EMBL; AF085275; AAD12055.1; -.
DR HSSP; P25685; 1HDJ.
DR MENDEL; 36329; Hevbr:1253; 36329.
DR PROSITE; PS00637; DNUJ_1; 1.
DR PROSITE; PS00637; DNUJ_CXKXG; 1.
KW Chapterone; DNA replication.
SQ SEQUENCE 415 AA; 46072 MW; C5F7C58A CRC32;

Query Match 27.0%; Score 508; DB 10; Length 415;
Best Local Similarity 35.9%; Pred. No. 8.5e-34;
Matches 128; Conservative 63; Mismatches 126; Indels 40; Gaps 10;

QY 26 FYKILGVPASIKDKKAYRKALQLHPDRNPDPQAOEKFODLAGAYEVLSDSEKRO 85
Db 14 YVELIGVKNMSODLLKAYRKALIKNHPDKGD---PEKRELAQAYEVLSDSEKREI 69
QY 86 YDTYEEGLKDGHOSSHG-----DIFSHFPGDFGFMGCTPRQODRNIPRGSIIYDLEV 140
Db 70 YDOYEDALKEGMSGGGGAGHDPFDIFOSFFGNGPFGGGSSGRKK---EGEDVLIHPLKV 126
QY 141 TLEEVYAGNF--VEVVRN-----KVARQAGKRCN-CR---QEMRTQLGPRFQWT 188
Db 127 SLEDLYNGTSKSLSTRNVYCSKCKGSKSASMSKCCSGGSKGVYSTRQLGPFMIQ 186
QY 189 Q-----EVV-----CDECPNVKLVNEERTLEVEIPGVDRGMEYEPFGESEPHVD 233
Db 187 QHPCKEKGSTGETINDKRCRCPGCKGEKVEYKGVLEIVEKGMQNGORTTFPGADAPD 246
QY 224 GEPGLFRIRKVVKHPFERRGDDLYTNVTVSLVESLVGFEMDITHDGHKVIISR-DKI 292
Db 247 TTGTGDIYVLOQKEHPKFRKKGDDLIYDHTLSLEALCASQFILTHTLGDLLIKSQPGEV 306
QY 293 TRPGAXXMKKGESLPFNDNNIKGSLITFDVDFPEKQLEEARRGISQLKQGSVO 349
Db 307 VKPDQKAINDEGMWYQRPFRKGLYIHFSVDFPDSLPPDCKALEAVLPSRTSVQ 363

RESULT 11
ID 024074 PRELIMINARY; PRT; 423 AA.
AC 024074;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DE 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE DNUJ-LIKE PROTEIN.
GN MS11.
OS Medicago sativa (Alfalfa).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons;
OC core eudicots; Rosidae; eustosids I; Fabales; Fabaceae; Papilionoideae;
OC Medicago.
```

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV, RANGELANDER;
 RA FRUGIS G., MELE G., GIANNINO D., MARIOTTI D.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA FRUGIS G., MELE G., GIANNINO D., MARIOTTI D.;
 RT "Isolation and characterization of a DnaJ-like gene from alfalfa."
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE
 CC PROTEINS.
 DR EMBL: AJ000995; CA04447.1; -.
 DR EMBL: AF069507; AAC19391.1; -.
 DR HSSP: P25685; 1HDJ.
 DR MENDEL: 26909; Medea; 1253; 26909.
 DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 KM Chapterone: DNA replication.
 SQ SEQUENCE 423 AA; 47197 MW; 7BB3C880 CRC32;

Query Match 27.0%; Score 507.5; DB 10; Length 423;
 Best Local Similarity 35.8%; Pred. No. 9, 6e-34;
 Matches 121; Conservative 57; Mismatches 117; Indels 43; Gaps 9;

QY 26 FYKILGPPRSASITDIKAYRKALALQHPDRNPDDPQAOEKFDGLGAAYVLSDSERKQ 85
 DB 15 YVDLGVSKASASDEIKAYRKAKMKNHPDKGDP---PEKFKELGAYVLSPEKKEL 70
 QY 86 YDFYEGELKDGHOSSHG-----DIFSHFGDFGFMGTFRQDORNPISGDIYVDL 138
 DB 71 YDQYGEALKEHMGKGSGSFHNPDIFFESFG-AGFGGGGPRSA--RROKQGEVHVHST 127
 QY 139 EYVLEEVYAGNFVEY-VRNKPVARAQAPGK-----RKCNCR---QEMRTTOLGPRFQ 186
 DB 128 KVSLEEDVYNGTTRKLSLRNALCSKCKSGSKGTAGRCFGCGTGKIRRDIGLGMIQ 167
 QY 187 MDEYV-----CDECPNVKLVNEERTLEVITEPGVYRDGMETPRIGSEPH 231
 DB 188 OMQHVCPDCKGTGEVISERDCPOCKGNKITOEKYLEVHEVKGMOQGHKIVFEGQDEL 247
 QY 232 VDGEPPGLRPRIRIKVKKPIPERGGDLYTNVTVSLVESLVGFEMDITHLGHVHLSRD- 291
 DB 248 PDITIGDLYVLOVKGHPKRRERDDLIHNNLSLTDALCGFQFNVTIHLDRQLLVKSNP 307
 QY 291 -KITRPGAXXMKKGGELPNFDNNNKGSLIITFDVDFP 327
 DB 308 GEVIRKPGQHNKAINDGMPQHGPRFMKGRILYKFSVDFP 345

RESULT 12
 ID 016303 PRELIMINARY; PRT; 439 AA.
 AC 016303;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
 DE T05C3.5 PROTEIN.
 GN T05C3.5.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 CC Rhabditina; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RX MEDLINE: 94150718.
 RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BRKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,

RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HARKINS T., HILLIER L., JER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
 RA LIGHTWING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SVALDON N., SMITH A., SONNHAMMER E., STADEN R., SUSTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLMAN P.,
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA BLANCHARD M., BRADSHAW H., KRAMER J.;
 RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA WATERSTON R.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1 SIMILARITY: TO OTHER PROKARYOTIC DNAJ, AND TO EUKARYOTIC DNAJ-LIKE
 CC PROTEINS.
 DR EMBL: AF016428; AAB65361.1; -.
 DR HSSP: P25685; 1HDJ.
 DR PROSITE: PS00637; DNAJ_CXXCXGXG; 1.
 DR PFAM: PF00226; DnaJ; 1.
 DR PFAM: PF01556; DnaJ_C; 1.
 DR PFAM: PF00684; DnaJ_CXXCXGXG; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 KM Chapterone: DNA replication.
 SQ SEQUENCE 439 AA; 47514 MW; B3453315 CRC32;

Query Match 26.3%; Score 495; DB 5; Length 439;
 Best Local Similarity 35.0%; Pred. No. 1, 1e-32;
 Matches 129; Conservative 56; Mismatches 116; Indels 68; Gaps 14;

QY 27 YKILGPPRSASITDIKAYRKALALQHPDRNPDDPQAOEKFDGLGAAYVLSDSERKQ 86
 DB 15 YTLNVRPDAQSADIKSYFKLAKETHPNKPDH---GDKFEISFAYVLSPEKKRLY 71
 QY 87 DTYGEGELKDGHOSS-----HGDIFSHFG-----DFGFMFGT----- 121
 DB 72 DARGLGVGGGAGGGGGGPGGLFHFEGAGGDDDDDDMGHPRFGGLGNGKMR 131
 QY 121 --PRODRNIPKSGDIYVLEVTLEVYAGNFVEY-VRNKPVARAQAPG-----KRCN 170
 DB 132 GGFRRR---KFQDTVHPLNVTLELTVYGTSKLSKALKCTCEGSGGKRGKRYCD 186
 QY 171 -CR-QEMRT--TOLGPRFOMTOEVVCECP-----NKLTVNEERTLEV 210
 DB 187 ACRGRGVKTIYQIIGFMGLQMO-VHCDACKSGSKVPAQDKCKCKGHEKVENVSILTEV 245
 QY 211 EIEPGVYRDGMETPRIGEG-EPHVDEPPGLRPRIRIKVKKPIPERGGDLYTNVTVSLVES 269
 DB 246 HVLPGKHNDKITFGKGDQSDPDGPDGVVVIQKHQDIDIRKRGDDLIHMKKSLMEWA 305
 QY 270 LVGFEMDITHLGHVHLS--RDKITRPGAXXMKKGGELPNFDNNNKGSLIITFDVDFP 327
 DB 306 LCGYVLELTKHLDGHPVLSKGGDVAKPQVIRGVILGKGMKPKKYPDELKGNLFEVEVEFP 365
 QY 328 KEQLTEAR 336
 DB 366 KEHFLDDEK 374

RESULT 13
 ID 066921 PRELIMINARY; PRT; 376 AA.
 AC 066921;
 DT 01-AUG-1998 (TREMBLrel. 07, Created)
 DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)

DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE CHAPERONE DNAJ.
 GN DNAJ.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5:
 RX MEDLINE: 98196666.
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus.";
 RL Nature 392:353-358(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5:
 RA DECKERT G., WARREN P.V., GAASTERLAND T., YOUNG W.G., LENOX A.L.,
 RA GRAHAM D.E., OVERBECK R., SNEAD M.A., KELLER M., ADJAY M., HUBER R.,
 RA FELDMAN R.A., SHORT J.M., OLSON G.J., SWANSON R.V.;
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: A6000703; AAC06881.1; -
 DR HSSP: P08622; 1XBL.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF01556; DnaJ_C1.
 DR PFAM: PF00684; DnaJ_CXXGXG; 1.
 DR PRINTS: PR00625; DNAJPROTEIN.
 DR SEQUENCE 376 AA; 41963 MW; 07BC58EB CRC32;

Query Match 26.0%; Score 490; DB 2; Length 376;
 Best Local Similarity 34.5%; Pred. No. 2.2e-32;
 Matches 128; Conservative 61; Mismatches 118; Indels 64; Gaps 15;

QY 24 RDYKILGVRSASIKDKKAYRKRLALQHPDNPPDPOAQEKRODLAGAYEVLSDSEK 83
 DB 7 KDYELIGVPRMSOEIKKAYRKRLKHPD-ICKKPECEKEKELNEAYQVLSDEKR 65
 QY 84 KQYDTYGEELKDGHOSSHG-----DIFSHFPG-DEGFMF---GGTPRQ 123
 DB 66 KLVDMV-----GHAFAEAGAAQARVETTELPIREILNEFFDFDGLSIFERATGRRA 118
 QY 124 QDNRTPGSDIIVDELYTEVTAAGNFVEYVRNKPVARQAPG-----KRC-NCROE 174
 DB 119 RRRRSYKGEIIVVPEITLEAFKGTVPPIEVEREYPCSAAGGTGDESKSRCTCGGR 178
 QY 175 MRTTQGGPRGFMTOGVVDCPCNVKLVNE-----ERTLEVELEPRVRCOME 221
 DB 179 GEIVQ-GNMFQVRO--TCPTCGEGEVIYENCHACTGGRGYGLVKEITKVKIPPGVRGSK 235
 QY 222 YPTIGDEEPV---DGEPRGLRFRIKVKVKKHPIFERRGDDLYTNVTVSLVSLVFEEMDT 278
 DB 236 --LVESGKAGAGYGGPRGLYIIVKKPKIKFERKGDLLVYNVITTPYPAVLSTEVEVP 293
 QY 279 HLDGKHVHISDKITRPGAXXMKKGEGLPNFDNNNIKSLITTFVDVDPK---EQLTTEE 334
 DB 294 TLDEGKKVXIPPGTEGELIKVPGKMPRLKSG--RGDLVVRVHIDVPRKIGVLSKILGD 352
 QY 335 ARGIGKOLLKO 345
 DB 353 GKK-VEELLKO 362

RESULT 14
 ID 020774 PRELIMINARY; PRT; 331 AA.
 AC 020774;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE F54D5.8 PROTEIN.

GN F54D5.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditidae;
 OC Rhabditina; Rhabditidae; Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA COLES L., MATTHEWS L.;
 RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
 RA BONFIELD J., BURTON J., CONNELL M., CORSEY T., COOPER J., COULSON A.,
 RA CRAXTON M., DEAR S., DU 2., DURBIN R., FAYELLO A., FULTON L.,
 RA GARDNER A., GREEN P., HAKINS T., HILLIER L., JIER M., JOHNSTON L.,
 RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LAVREILLE P.,
 RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
 RA PARSONS J., PERCY C., RIEKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
 RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., STULTON J.,
 RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
 RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 DR EMBL: Z66513; CA91334.1; -
 DR HSSP: P25685; 1HDU.
 DR PROSITE: PS00636; DNAJ_1; 1.
 DR PFAM: PF00226; DnaJ_1.
 DR PFAM: PF01556; DnaJ_C1.
 DR SEQUENCE 331 AA; 36265 MW; E00F46D5 CRC32;

Query Match 25.0%; Score 470.5; DB 5; Length 331;
 Best Local Similarity 32.4%; Pred. No. 7.3e-31;
 Matches 119; Conservative 55; Mismatches 102; Indels 91; Gaps 9;

QY 23 GRDYKILGVRSASIKDKKAYRKRLALQHPDNPPDPOAQEKRODLAGAYEVLSDSEK 82
 DB 2 GKDYKVLGSKGWTDEIKKAYRKRLKHPDN-KEAGENKFKELAEAYDVLSDDK 60
 QY 83 KQYDTYGEELKDG-----HOSSHG--IFSHFPG----- 113
 DB 61 KTIYDQGEELKDGKGGAGGGGGMHYFERGDPNMFSSFFGSDPFGAGGEMDGL 120
 QY 113 -----FGFMGTPRQDNRDIPGSDIIVDELYTEVTAAGNFVE 152
 DB 121 GGAGGPMFMFNOGMDMGFMGMOGGRGHARQDPA---VLHDLVSLEEDVLKG--- 174
 QY 153 VVRNKPVARQAPGRKCKNCQKQEMTTLQGRGFMTOGVVDCPCNVKLVEERTLEVEI 212
 DB 174 -----TTR---KMKITRKVMDNQR---EDKVLVTI 201
 QY 213 EPGVROGMEYFIEGEPHYDGEPRGLRFRIKVKVKKHPIFERRGDDLYTNVTVSLVSLVFE 272
 DB 202 KPGKSGTKITFRPESGQHPRTADIVYIKDPRHFKFKESGDIKRVKISLSKALTG 261
 QY 273 FEMDTHLDGKHVHISDKITRPGAXXMKKGEGLPNFDNNNIKSLITTFVDVDPKQDLT 332
 DB 262 LDIMPTLDGADVRLQNDVYIKPGTRRLATGKLPNPKPSHNRDILLTFEVEPR-SQLN 320
 QY 333 EAREGI 339
 DB 321 PTOREVI 327

RESULT 15
 ID 092FC5 PRELIMINARY; PRT; 371 AA.
 AC 092FC5;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, last sequence update)
 DT 01-NOV-1999 (Tremblrel. 12, last annotation update)
 DE PUTATIVE DNAJ.

